日本語音声認識システムに関する研究(Dissertation_全文)

AUTHOR(S):

中津, 良平

CITATION:

中津, 良平. 日本語音声認識システムに関する研究. 京都大学, 1982, 工学博士

ISSUE DATE:

1982-11-24

URL:

https://doi.org/10.14989/doctor.r4815

日本語音声認識システムに関する研究

昭和57年4月

中津良平

日本語音声認識システムに関する研究

中　津　良　平

日本語音声認識システムに関する研究

中　津　良　平

本論文は，音声による人間と機械の対話の実現を目ざして行った日本語音声認識システムに関する一連の研究の総合報告である。

認識の対象としては，日本語音声におけるほぼすべての対象を取り扱う目的で，単語音声，連続単語音声，会話音声を取り上げる。また，認識の基本的な手法としては，母音一子音一母音よりなるＶＣＶ音節を単位とする方法と，音韻を単位とする方法を取り上げる。これらの方法を各対象に適用した場合の認識方法の詳細について述べると共に，得られる性能を明らかにする。さらに，音声入力装置の実現を目ざし，単語音声，連続単語音声，会話音声を対象とした認識装置，オンライン認識システムを作成し，評価実験によりその性能を明らかにする。

目　　次

主　な　記　号　の　表

記　　号	意　　味	備　　考
w	角周波数	
p	分析の次数	
$T(w)$	最尤スペクトル分析によるスペクトル包絡（標準パターンを表現するのに用いる）	
u_i	自己相関関数（標準パターンを表現するのに用いる）	
α_i	線形予測パラメータ（標準パターンを表現するのに用いる）	
A_i	最尤スペクトルパラメータ（標準パターンを表現するのに用いる）	
σ_T^2	残差パワー（標準パターンを表現するのに用いる）	
$R(w)$	最尤スペクトル分析によるスペクトル包絡（入力音声を表現するのに用いる）	
v_i	自己相関関数（入力音声を表現するのに用いる）	
ρ_i	自己相関係数（入力音声を表現するのに用いる）	
β_i	線形予測パラメータ（入力音声を表現するのに用いる）	
B_i	最尤スペクトルパラメータ（入力音声を表現するのに用いる）	
σ_R^2	残差パワー（入力音声を表現するのに用いる）	
$\boldsymbol{V}=(v_1, v_2, \cdots\cdots, v_N)$	入力音声の自己相関関数による表現	
$\boldsymbol{P}=(\rho_1, \rho_2, \cdots\cdots, \rho_N)$	入力音声の自己相関係数による表現	
N	入力音声のフレーム数（1フレーム＝15 ms）	
$\boldsymbol{v}_i=(v_{i0}, v_{i1}, \cdots, v_{ip})$	入力音声の第 i フレームの自己相関関数	
$\tilde{\boldsymbol{V}}=(v_N, v_{N-1}, \cdots, v_1)$	入力音声の時間軸を逆にした系列	10章
$\boldsymbol{\rho}_i=(\rho_{i0}, \rho_{i1}, \cdots, \rho_{ip})$	第 i フレームの自己相関係数	

主　な　記　号　の　表（続き）

記　　号	意　　味	備　　考
$\boldsymbol{R}=(r_1, r_2, \cdots\cdots, r_M)$	標準パターンの自己相関係数による表現	
$(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots\cdots, \boldsymbol{A}_M)$	標準パターンの最尤スペクトルパラメータによる表現	
M	標準パターンのフレーム数	
$A_j=(A_{j0}, A_{j1}, \cdots, A_{jp})$	第 j フレームの標準パターンの最尤スペクトルパラメータ	
$(A_{x0}, A_{x1}, \cdots\cdots, A_{xp})$	音韻 / x / の標準パターン	
$\boldsymbol{Q}=(\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots, \boldsymbol{q}_M)$	複数の学習サンプルより標準パターンを作る際の基準サンプル	3 章
$P=(v_{10}, v_{20}, \cdots, v_{N0})$	入力音声のパワーの系列	
$\tilde{P}=(\tilde{v}_1, \tilde{v}_2, \cdots\cdots, \tilde{v}_{N-1})$	平滑化された音声パワーの系列	4 章
$(e_1^1, e_2^1, \cdots\cdots\cdots, e_N^1)$	急激なスペクトルの変化を示す系列	5 章，7 章
$(e_1^2, e_2^2, \cdots\cdots\cdots, e_N^2)$	ゆるやかなスペクトルの変化を示す系列	5 章，7 章
$P_1,\ P_2,\ P_3$	音声検出に用いる閾値	
P_T	セグメンテーションの閾値（無音区間検出用）	
Q_T	セグメンテーションの閾値（パワーの谷部検出用）	
$dv_1,\ dv_2$	音声パワーの極小値と前後の極大値との差	
$E_1\ ,\ E_2$	スペクトル変化を用いたセグメンテーションの際用いる閾値	5 章
$\{S_k^n\}, \{S_k^M\}\ (k=1, 2, \cdots)$	セグメンテーションの際用いる閾値	4 章
$\{G_1, G_2, \cdots\cdots\ \}$	セグメント境界の候補（各 G_i はセグメント境界のフレーム番号の集合）	4 章
$\{C_1, C_2, \cdots\cdots\ \}$	〃　（各 C_i　〃　）	5 章
d_{ij}	第 i フレームのスペクトルと第 j フレームのスペクトルの差	

主　な　記　号　の　表（続き）

記　　号	意　　味	備　　考
$l(\boldsymbol{\rho}_i, \boldsymbol{q}_j)$	特徴パラメータ $\boldsymbol{\rho}_i$ と $\boldsymbol{q}_j$ の間の類似度	
$l(i, j)$	入力音声の第 i フレームと標準パターンの第 j フレームの間の類似度	
$l(i, x)$	入力音声の第 i フレームと音韻 $/x/$ の標準パターンとの類似度	
$LM=\{l(i, j)\}$	類似度マトリクス	
f	入力音声と標準パターンの非線形な対応づけを行う関数	
$f(i)$	入力の第 i フレームに対応づけられる標準パターンのフレーム番号	
w w^r $(r=1, 2, \cdots\cdots, R)$	} 認識対象の単語	
R	認識対象の単語数	
$x_1\, x_2 \cdots\cdots x_j \cdots\cdots x_J$	w^r の音韻系列表示	9章，10章，11章
D_j	x_j の平均継続時間	〃
D_j^m	x_j の最小継続時間	〃
D_j^M	x_j の最大継続時間	〃
$wa^r= y_1 \cdots\cdots y_k \cdots\cdots y_K$	w^r を展開した音韻系列	〃
$wb^r= y_1 \cdots\cdots y_k \cdots\cdots y_K$	〃	〃
$\widetilde{w^r}$，$\widetilde{wb^r}$	w^r，wb^r の時間軸を逆にした音韻系列	10章
$L(\boldsymbol{P}, \boldsymbol{Q})$	パターン $\boldsymbol{P}$ とパターン $\boldsymbol{Q}$ の間の類似度	
L	入力音声と標準パターンの間の類似度	
$L(r)$	入力音声と単語 w^r の類似度	

主　な　記　号　の　表（続き）

記　　号	意　　味	備　　考
$L(i, j \mid w)$	入力音声の部分系列（第iセグメント～第jセグメント）と単語wの類似度	
$L(i, j \mid r)$	入力音声の部分系列（1～iフレーム）と単語w^rの部分系列（1～jフレーム）の類似度	
$L(i, j)$	$L(i, j \mid w)$のwに関する最大値	
$w(i, j)$	$L(i, j)$を与える単語	
L_j	入力の部分系列（1～jセグメント）と連続単語の間の最大類似度	
$LA(i, j \mid r)$	入力音声の部分系列（i～jフレーム）と単語w^rの類似度	10章
$LA(i, j)$	$LA(i, j \mid r)$のw^rに関する最大値	〃
$w(i, j)$	$LA(i, j)$を与える単語	〃
LA	入力音声と連続単語の最大類似度	〃
WA	LAを与える単語系列	〃
$LB(i)$	入力音声の部分系列（1～iフレーム）と連続単語の最大類似度	〃
$WB(i)$	$LB(i)$を与える単語系列	〃
$LC(i, j \mid r)$	入力の部分系列（1～iフレーム）と単語w^rの部分系列（1～jフレーム）の類似度	〃
$LC(i)$	入力の部分系列（1～iフレーム）と標準パターンの最大類似度	〃
$WC(i)$	$LC(i)$を与える標準パターン	〃
$\widetilde{LC}(i, j \mid r)$	入力$\widetilde{V}$の部分系列（1～iフレーム）と単語$\widetilde{w}^r$の部分系列（1～jフレーム）の類似度	〃
$LD(i \mid r)$	入力音声の第iフレームにおける単語w^rとの類似度	〃

第1章　序　　論

1.1　音声認識

人間が社会生活を行っていく上で，お互いの意志疎通，情報伝達は欠くことができないものである。情報伝達の手段としては，音声，身ぶり，文字，図形等があるが，中でも音声は，手軽であり，かつ広範囲に使用できる情報伝達手段として日常生活で最も広く使われているものである。この音声を，人間同士の情報伝達手段としてばかりでなく，人間と電子計算機等の機械との間の情報伝達手段として使おうとするのが音声認識研究の目的である。

我々が音声を使って情報を伝える手順は以下のようになる。まず，頭の中にある，相手に伝えたい考え，意図を単語の系列である文章に変換する。この際，使用している言語，話し手の知っている語彙等にしたがって構文，単語が決定される。次に文章は音韻の系列に変換される。この時，イントネーション，アクセントといった韻律情報が加えられる。音韻の系列にしたがって発声器官を動作させるための神経パルスが発生され，これによって発声器官が動作し，音声波が生成される。発声器官は質量を持っており，慣性のため不連続的な動作は出来ないから音韻系列というディジタル量はここで連続的なアナログ量に変化することになる。これは音声波レベルで見ると，各音韻に対応した音声波が互いに影響を及ぼし合って変化することをさしており，調音結合と呼ばれる。さらに，上に述べた，意図から音声波への変換の各段階で，相手に伝えたいと思っている感情が物理量に変化して加えられる。これは，例えば意図から文章への変換時には，平叙文を使うか感嘆文を使うかといった差であらわれ，文章から音韻系列への変換時には，イントネーションのつけ方，アクセントの位置，強弱の差となってあらわれる。また，話し手の個人差も付加される。これは，発声器官の個人差のために発声器官の動作の際に加えられるのはもちろんであるが，他にも，使用する語彙，方言の差などにより各段階で付加される。聞き手は，音声波を耳で受け取り，聴覚器官で神経パルスに変換し，脳の中で音韻系列，文章へと逆変換してゆき，最後に，話し手の伝えようとした意図を抽出する。当然，この認識の各段階で，話し手の感情，個人性といった情報も抽出され，聞き手に了解される。以上の過程を図1.1に示す。

以上の説明から明らかなように，音声には次の3種類の情報が含まれていると考えてよい。

(1)　音韻情報……話し手が伝えようとしている意味内容は何か。

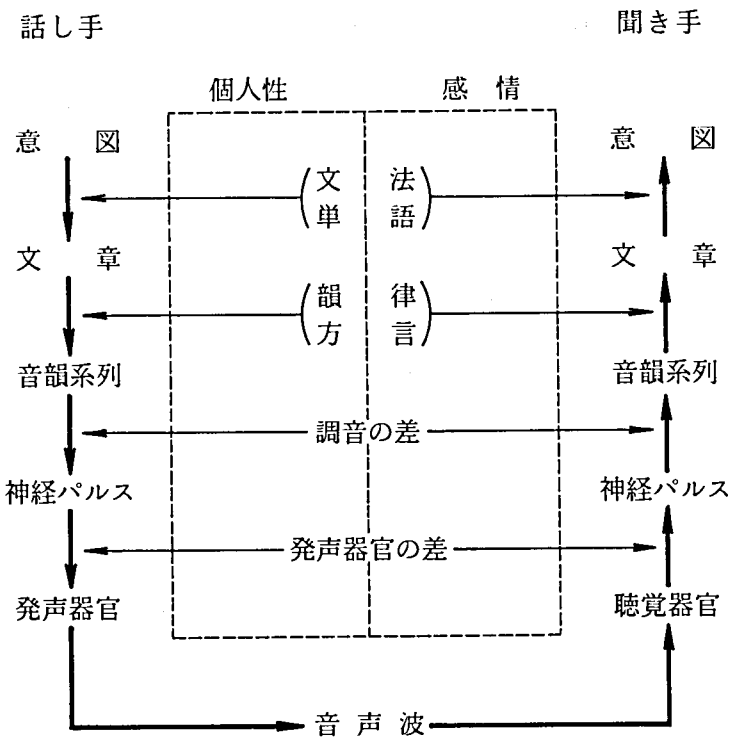


図 1.1　音声の発声・認識過程

(2)　個人性情報……話し手は誰か。

(3)　情緒情報……話し手はどのような感情を持っているか。

音声認識は広義には音声に含まれる(1),(2),(3)の情報をぬき出すことであるが，通常使われる狭義の意味では(1)をさす。本論文では，(1)の音韻情報のみを取り扱うことにする。

さて，話し手と聞き手の間では，上に述べたような複雑な過程を経て音声による情報伝達が行われるわけであるが，人間の脳の中における変換の具体的なプロセスについては，現在のところ十分な知識は得られていない。そのため，音声の自動認識の研究は試行錯誤的に研究を進めて行かねばならず，困難な面が多い。音声認識を困難にしている点を整理すると以下のようになる。

(1)　認識対象

(2)　言語情報

(3)　調音結合

(4)　個人差

ここでは感情といった高次の情報は除いて考えることにする。まず(1)は，使われている言語は日本語か否か，対象は一般的な会話音声か，それとも区切って発声された単語音声か，また

記彙は数十か数百か，それとも制限はないのかといった音声認識の対象の範囲を設定する要因をさす。この設定の範囲の広さによって音声認識の困難さが数段ちがってくる。次に，(2)の言語情報とは，我々が文章をしゃべる時に単語を文法に従って並べるが，その際の法則をさす。これは当然言語によって異ってくるし，しかも，文法という形できちんと整理するのが困難であり，例外の多い規則にならざるを得ないという難点がある。さらに困難なことは，会話においては文法通りに発声されない事が多いという点である。(3)の調音結合は，先に述べたように発声器官の慣性のために，音韻系列に従って生成された音声波が不連続ではなく，ある音から次の音へ連続的に変ることを示している。このことは，認識する立場から見ると，音声中の各音韻が互いに影響し合って変形しているため，切り出して同定することが困難であることを示している。最後の(4)の個人差は，発声器官の差，話し方の差などの先天的，後天的な個人差により同じ意味内容をもっている音声でもその物理的性質が個人によって異なることを示している。

さてそれでは，以上のような困難な問題を持った音声認識が実用的には意味をもたないかというと，そのようなことはない。音声を人間と電子計算機等の機械との間の情報伝達手段として使うことによる利点を実用的立場から見ると以下のようになる。

(1) 人間にとって最も自然であり，便利である入力手段であり，特別の訓練をしなくても誰でも容易に使用できる。

(2) 情報の発生を高速度に行うことができる。表1.1に種々の手段による情報伝達の速度を示すが，音声は文字を書いたりタイプライタを打鍵したりする場合に比較して，2.5～6倍の速度が得られる。

表1.1 種々の入力形式の情報速度

入力形式	情報速度（語/秒）
音声入力 朗読	4
自然な発話	2.5
孤立発声	1
キーボード入力（熟練者）	1 （5ストローク/秒）
オンライン手書き文字	0.4
押ボタン入力	0.3（1.5数字/秒）
マークシート	0.1（0.5数字/秒）

(3) 手がふさがっていても自由に入力できる，情報入力源（マイク）からはなれても良い，歩きまわりながら情報入力できるなど，他の入力手段にない特徴を持っている。

(4) 認識結果をすぐ話し手にフィードバックして確認，訂正を求めるといったオンライン的な使い方に向いている。

したがって，認識対象を100語程度の単語に限定することで認識対象，言語情報の問題をさけ，単語単位の認識を行うことで調音結合の問題をさけ，さらに，利用者ごとに標準パターンを登録しておくという方式で個人差の問題をさけることにより，一部では実用化されつつある。今後利用者が広がってゆく分野として次のようなものが考えられる。

(1) 製品検査等のデータの音声入力。

(2) 荷物等の区分を音声で行う。

(3) 座席予約，情報検索の際の音声入力。

(4) 計算機端末への音声によるコマンド，データの入力。

さらに，個人差の問題に取り組み，不特定話者を対象とした音声認識研究も最近精力的に行われており，今後，電話機を入力端末とした各種サービスへ音声認識の応用が広まってゆく可能性がある。

1.2 音声認識の研究動向

音声認識は，単語音声を対象にした単語音声認識と，会話音声を対象にした会話音声認識に大別される。ここでは，文献(1)～(10)等を参考にしつつ単語音声認識と会話音声認識の研究動向を述べることにする。

単語音声認識のパイオニア的研究は，1958年のDavis等による数字音声認識に関する研究である[11]。そこでは，第1ホルマントと第2ホルマントの軌跡によって数字音声を認識しようとしている。そこで用いられている，単語全体を1つのパターンと見て認識を行う手法は，その後の単語音声認識方法の1つの主流になっている。また，帯域通過フィルタによって特徴パラメータを抽出する方式は，現在でも広く用いられている手法である。

その後，各国で単語音声認識研究がさかんに行われるようになり，各種の検討が行われた。それらについて以下に述べる。

まず，特徴抽出の方法は，その後，帯域通過フィルタの数を10～20に増やし，精密な周波数

スペクトルの抽出がはかれるようになった。[12]～[15] その後特徴抽出について特記すべき事項は，斎藤，板倉による最尤スペクトル分析，[16] PARCOR分析[17]の開発，および，ほぼ同時期にAtalらによって提案されたＬＰＣ分析[18]の開発である。これらはいずれも基本的には等価な方法であって，音声波の分析に時系列解析の手法を取り入れたものである。これらの手法はいずれもディジタル処理に向いており，折からの電子計算機の発達と結びつき，現在では音声の特徴抽出法の主流になっている。そして，これらの分析法によって得られる，自己相関係数，線形予測係数，PARCOR係数等が特徴パラメータとして広く使われるようになっている。[19]～[24]

次に認識の単位としては，Davisらと同じように単語を認識の単位とする方法もいくつか試みられた。[25][26] また，単語より小さい単位として，音声の基本的な単位である音韻を単位とする認識法が試みられた。[12][27][28][29] また，音韻より大きい単位として，音節等を単位とする方法も試みられた。[23][30][31][32] これらの手法は，音声を認識の単位に分割するセグメンテーションという手段が必要であり，その段階での誤りがさけられない。したがって，これらの方法は高い認識率を求めるというよりは，更に進んだ一般的な連続音声を対象とした音響処理の第１段階として単語を扱ったものが多い。一方，中間的な方法として，標準パターンとしては音韻等を単位としたものを用いるが，セグメンテーションを行わず，単語辞書を用いて単語全体で認識を行おうとする方法も試みられた。[19]

人が単語を発声する時，発声速度の変動により単語パターンの時間長が異なってくる。したがって，入力パターンと標準パターンのマッチングを行う際には，このような時間長の違いを補正してやる必要が生じる。初期の単語音声認識研究では，単に音声パターンの時間軸を線形に伸縮してやることによって，補正を行ってやろうとしていた。[14][25] しかしながら，このような方法では不十分であることが徐々にわかってきた。それは，たとえ同じような発声をしても，人間の発声にさけられない微妙なゆらぎのため，時間軸が非線形に伸縮しているためである。これは，発声者が異なれば当然生じるが，同じ発声者が同じ単語を同じ様に発声しても生じる問題である。この問題に対する解法として，ダイナミック・プログラミング法（Dynamic Programming；以下，本論文中ではＤＰと略記する）を用いた時間軸の非線形伸縮を伴ったマッチング法が提案された。これは，ほぼ同時期に，Velichko・Zagoruiko,[33] 迫江・千葉,[34] 好田[35]らによってなされた。その後，各種の研究により，この手法が極めて有効であることが明らかにされた。また各種のＤＰの比較実験も行われた。[36][37][38] 現在では，ＤＰは単語音声認識におけるマッチング法として定着しているのみならず，音声認識全般にわたる基本的なマッチング手法として広く用いられている。

単語の認識率を上げるためには，音響的な性質を使うだけでなく，言語的な知識を使うことが必要である。言語的情報としては，音韻間の遷移確率，調音規則，単語辞書等がある。また，単語認識系における音響レベルの処理の性質を利用する方法として，confusion matrixを用いる方法がある。これらの利用法とその効果について検討が行われた。[39][40][41]また，個々の単語がお互いに前後関係を持って発声される場合には，構文情報を利用することもできる。それについての検討も行われた。[42]〜[46]

単語音声認識で次に問題になるのは，発声者による声の質の違いである。Davisらによる実験では，標準パターンを登録した発声者に対しては認識率が良いが，そうでない発声者に対しては認識率が大幅に低下することが報告されている。この対策としてまずとられたのは，平均的な標準パターンを登録するなどの方法により普遍的な認識系を作成することであった。[47][48]しかしながら，平均的なパターンで代表するには個々の発声者の声が違いすぎることが明らかになってきた。これに対する実際的な解決策として，発声者ごとに学習により標準パターンを作成し，系を発声者に適応させる方法がとられた。[19][49]そして，学習の方法に関して各種の検討が行われた。[50][51]また，認識対象の語彙が多い場合に簡単に学習を行う方法についても検討された。[52][53]この，発声者ごとに標準パターンを登録する方式は，高い認識率を得る上で極めて実用的な方式であり，実用化されている認識装置の多くもこの方式によっている。このような方式は，一般に特定話者用の認識方式と呼ばれている。これに対し，不特定多数の発声者を対象とした認識方式は，不特定話者用の認識方式と呼ばれている。現在のところ特定話者用認識方式が主流であるが，最近では発声者による声の性質の差を正規化しようとする試みがいくつかなされている。[54][55]更に，語彙を比較的少数に絞り，最初から不特定話者用音声認識をねらいとする方式もいくつか発表されている。[56]〜[63]

次に認識対象であるが，Davisの研究に見られるように，初期の研究では，10数字を対象として取り扱うものが多かったが，徐々に拡大し，数百語を対象とする例もあらわれてきている。[34][64][65]認識された結果を1問1答形式で確認し，誤っていれば言い直すという方式も検討されている。[66][67][68]また，発声の形式としては，最初は区切って発声した単語を対象としていたが，数字等を連続して投入したいという要求から，連続して発声された単語音声を対象とすることが試みられた。[34][69]これについては，単語間のセグメンテーションが必要という困難な問題があるが，ＤＰを用いることによりセグメンテーションを必要とせず，連続単語を認識できる方式が提案され，[24][70][71]有効性が確かめられている。現在では連続数字等は十分実用レベルに達している。

また，最近のオフィス・オートメーションの発達に伴い，音声による日本語入力が音声認識

の応用分野として有望視されており，単音節認識等も試みられている。[72][73][74]

単語音声認識の研究を活発に行っている研究機関としては，アメリカでは，ベル研，IBM，BBN，RCA，MITなどで古くから研究が行われており，最近では，Threshold Technology社，Interstate Electronics社，Vervex社等のベンチャーメーカーが認識装置の製造，販売を行っている。その他の外国では，ソ連，イギリス，フランス，ドイツ，イタリア，カナダ，オランダ等の研究機関で行われている。また，日本では，京大，東大，東北大，北大，早大，通研，国際電々，電波研，日電，日立，富士通，松下電器等で研究されている。また，音声認識の有望さに注目して，各メーカーが続々と認識装置の開発，製造，販売を手がけはじめている。各研究機関で行われた単語音声認識研究の比較を表1.2に示す。

一方，会話音声認識のパイオニア的研究は，坂井・堂下[75]およびOlson[76]によるものである。これらの研究は，音響的な処理に重点をおき，いわゆる音声タイプライタの実現を目ざしたもので，極めて先駆的な研究であるが，同時に，音声現象の複雑さを示すことにもなり，会話音声認識に言語情報を使うことの必要性を研究者達に印象づけた。その後，言語情報を使った会話音声認識の研究がいくつか試みられたが，[77][78][79]いずれも対象を広く設定しており十分な結果は得られなかった。このことは，一般的な音声タイプライタの実現という問題設定自体に問題があるという考えをいだかせることになった。

一方，人工知能の分野で，自然言語処理研究が進み，自然言語を用いたいくつかの質問解答システムが作成された。[80][81]これらの自然言語処理の研究は，対象を限定すれば，従来は複雑で取り扱うのは不可能とされていた自然言語を扱えることを示した点で大きな意味がある。これらの研究に触発され，米国で1971年にARPA（Advanced Research Projects Agency）の援助による音声理解のプロジェクトが開始された。[8] 音声理解（Speech Understanding）という言葉は，従来の音声認識（Speech Recognition）に対比してこのプロジェクトで始めて使用された言葉であり，音声を構成している一語一語あるいは一音一音を正確に認識することではなく，発声された音声から，話し手が伝えようとした意味内容を抽出することに重点をおいていることが特徴である。ARPAの音声理解プロジェクトは，1976年末に仕様を満足するデモンストレーションシステムを作成するという目標のもとに研究が開始され，精力的に研究が行われた。[82]～[91] 当初は，米国の主な音声研究機関を含んでいたが，途中で，CMU，BBN，SRIとSDCの3つのグループに絞られた。会話の話題としては，チェスゲーム，[82]各種の情報検索[83][85][86][88][89]旅行の予算管理[87]などが取りあげられた。言語情報はトップダウン的に利

表 1.2　単語音声認識の研究

(a)　外国

研究機関	認識単語	発声者数 (内女性数)	認識のための 音声単位	学　習	認識率(%)	特　徴	年　代
ベル研	10 数字	1	単　語	有 (無)	97～99 (50～60)	F1—F2 平面の軌跡パターン，相互相関係数	1952
	10 数字	7 (1)	単　語	有 (無)	94 (67)	17ch —時間のスペクトル，相互相関係数	1960
	26 単語	13 (4)	音　韻	有 (無)	95 (72)	振幅等時間領域のパラメータ，Tree 規則	1968
	200 単語	1	単　語	有	97.3	LPC，Itakura 尺度，DP	1975
	10 数字 (3桁連続)	10 (5)	音　韻	無	91.0	LPC，零交差波，パワー	1976
	10 数字	110	単　語	無	98.2	LPC，マルチテンプレート	1979
	10 数字 (2～5桁連続)	6 (3)	単　語	無	97～99%	LPC，端点フリー，2段DP	1980
IBM	10 数字	50 (25)	音　韻	有 無	99 93	40ch，Tree 規則	1962
	15 単語	3 (1)	単　語	有 (無)	97.6～99.8 (54～85)	16ch —時間のスペクトル，線形判別関数	1966
	30 単語	1	単　語	有	98.7		
	16 単語	13 (2)	単　語	無	94.2	8 ch — 6セグメントのスペクトル	1971
BBN	54 単語	2	音　韻	有	94～96	19ch，44個の特徴	1968
	114 単語	1			96		
	54 単語	2			97	言語情報利用	1975

表 1.2　単語音声認識の研究（続き）

(a)　外国

研究機関	認識単語	発声者数（内女性数）	認識のための音声単位	学習	認識率(%)	特徴	年代
RCA	100単語		単語			8 ch，論理的組合せ	1961
	10数字			有	87.5		1967
MIT	54単語	10	音韻	無	86.3	16ch，ストレスの位置検出，その前後の特徴	1966
UNIVAC	100単語	5	音韻	無	94.0	273 Hzごとの BPF	1972
TTI	21単語	6	単語	有	99.5	19ch，32個の特徴	1972
	10数字	10	単語	有	99.8	VIP— 100	1974
	12単語	12	単語	有	97.2		
	15単語	20	単語	有	99.4		1976
	35単語	12	単語	有	98.5	Threshold 500	
	32単語	5	単語	有	99.7	Quick Talk（2段DPの変形）	1980
XEROX	91単語	1	単語	有	99.6	BPF（20ch）とLPCの比較，DP	1976
	36単語	1	単語	有	98.0		
ソ連	203単語	2	単語	有	94.8	5ch，DP（科学アカデミー）	1970
	150単語		音韻	無	90.0	（Lvov Univ.）	1976

表 1.2　単語音声認識の研究（続き）

(a)　外国

研究機関	認識単語	発声者数（内女性数）	認識のための音声単位	学習	認識率(%)	特徴	年代
英	10数字	9 (2)		有 (無)	93 (59)	自己相関関係，パターンマッチング	1968
	10数字	60 (30) 24 (12)		有 無	94.3(男) 91.0(女) 95.8(男) 88.3(女)	零交差情報	1970
伊	10数字	10	音韻	無	70～90	17 ch，音韻の出現順序	1967
	10数字	10	単語	無	93.6	10 ch，組合せ順序回路	1970
独	14単語	10	単語	有 無	97 87	10 ch－3セグメントのスペクトル，論理式	1965
加	15単語	14 (7)	音韻	有	96.8	波形の非対称性を利用，Tree規則	1968
仏	36単語	1 2	単語	有 無	99 94	24 ch，構文情報利用，DP	1974

表 1.2　単語音声認識の研究（続き）

(b)　国内

研究機関	認識単語	発声者数（内女性数）	認識のための音声単位	学習	認識率(%)	特徴	年代
日電	10数字	1 20	音韻	有 無	99.7 97.9	8個の特徴，ベーズ則，ハード化	1964
	10数字	4	単語	有	99.8	16ch BPF，DP	1971
	10数字（4桁連続）	5	単語	有	99.2	12ch BPF，2段DP	1976
	100単語	40 (3)	単語	無	99.1	カテゴリ毎に特徴点設定，区分的線形判別関数	1977
	45音節	5	音節	有	98.1	FFT，DP，マルチテンプレート	1979
日立	10数字	1	単語	有	98	PARCOR係数，DP	1972
富士通	100単語	2 (1)	単語	有	93～99	14ch BPF，時間軸を11分割，ユークリッド距離	1969
	10数字	5	音韻	有	92.8	PARCOR係数	1973
	10数字	135	音韻	無	97	13ch BPF，細分類音韻パターン	1980
	1016文節	5	文節	有	95.8	LPC，ルート限定マッチング	1981
京大	10数字	40	音韻	有 無	98.4 95.8	部分学習	1976
	100単語	5	音韻	有 無	93.3 83.6	大局・局所的特徴による前照合	1978
	10数字	30	単語	無	97.6	20ch BPF，時間軸・周波数軸・強度軸上の伸縮	1980

表 1.2　単語音声認識の研究（続き）

(b)　国内

研究機関	認識単語	発声者数 (内女性数)	認識のための 音声単位	学　習	認識率(%)	特　徴	年　代
東　大	30単語	1	単　語	有	99.6	15ch BPF，不均一標本化，DP	1977
東北大	20単語	5	音　韻	無	86～96	29ch BPF，ローカルピーク	1976
	166単語	3	音　韻	無	82.0		
	51単語	23	音　韻	無	95.1	29ch BPF，ローカルピーク，confusion matrix利用	1980
	166単語	23	音　韻	無	87.7		
早　大	11数字	1	音　韻	無	94 2	音韻推移行列	1971
国際電々	10数字	1 10	音　韻	有 無	97.7 70	18ch BPF，書換え規則	1972
電波研	10数字	7	音　韻	有 無	96 80	26ch BPF，重みつきハミング距離	1962
NHK	10数字	13			95		1973
通　研	10数字	68 (30)	音　韻	有	99.8 (男) 98.6 (女)	最尤スペクトル分析，DP	1972
	90単語	8	VCV音節	有	97.0	最尤スペクトル分析，DP，VCV単位の認識	1973
	158単語	12	音・韻	有	97.6	質問回答形式	1978
	10数字 (1～4桁連続)	11	音　韻	有	99.7	LPC，2段DP，逆DP	1978
	641単語	1	音　節	有	95.0	LPC，クラスタ化した音韻標準パターン	1980

表 1.2　単語音声認識の研究（続き）

(b)　国内

研究機関	認識単語	発声者数（内女性数）	認識のための音声単位	学習	認識率(%)	特徴	年代
通研	101音節	4	音節	有	93.3	LPC，子音大分類，DP	1981
	16単語	620	音韻	無	96.1	LPC，マルチテンプレート，DP	1981
東芝	68音節	5	音節	有	95.2	16ch BPF，逆時間DP	1980
沖	12単語	130	単語	無	98.9	22ch BPF，選択的重みづけマッチング	1981

表 1.3　ARPAの音声理解プロジェクトの目標

音声の種類	連続音声
発声者の数	多数
発声者の種類	標準のアメリカ英語を話す協力的な発声者
発声環境	静かな部屋
変換器	高品質マイクロホン
発声者への適応	若干の調整
発声者の訓練	自然な適応
語彙数	若干選択された 1,000 単語
言語の種類	人工的文法
タスクの種類	データマネジメント，計算機状態の問い合せなど
発声者の心理モデル	単純な心理モデル
対話の複雑さ	自然なやりとり
許容できる誤り	10％以下の意味の誤り
処理時間	実時間の数倍
稼動年月	1976 年

用することが多いが，並列的に利用することも検討された。[82][83]　構文解析の手法は，通常の depth-first method や breadth-first method の他にそれらの中間的な方法にあたる best-first method[83][87]も試みられた。1976 年末にプロジェクトが終了したとき目標に近い性能を示したのは CMU の Harpy システム[85]のみであった。Harpy システムは，言語情報を推移網で表現しており，推移網中には，すべての可能な構文，単語のすべての可能な音韻表記，単語境界におけるすべての可能な調音規則が埋め込まれている。認識方法はこの推移網から入力音声に最もよくマッチするパスを発見することであって本質的には単語認識と同じ処理を行っていることになる。このように，比較的単純な構成のシステムが最も良い成績を上げたということは，問題が残る。したがって，言語情報の高度な利用により，実用にたえる会話音声認識システムを作り上げるというARPAの最初の目標は十分に目的を達しないまま終了した。その主な原因は音響レベルの研究が弱かったことにあると指摘されている。[9][10]しかしながら，音声認識における，構文，意味，プラグマティクスなどの高度の言語情報の表現方法，利用方法が具体的になってきた点は大きな意義がある。

ARPAの音声理解プロジェクトは各国の研究機関に大きな刺激を与えた。日本でも通研[92][93][94]，京大[95]，京工繊大[96]，山梨大[97]などで研究が開発された。米国では，ARPAのプロジェクトとは別にＩＢＭなどで精力的に研究が行われた[98]。ヨーロッパにおいても，1976年ごろから，フランス，イタリアなどで開始された[99]～[101]。ARPAのプロジェクトが終了した後，米国では会話音声認識は下火になったが，ＩＢＭ[102]等では地道に研究が続けられている。またベル研でも1979年から研究が開始されている[103]。また，日本やヨーロッパではむしろ米国より熱心に研究が続けられている[108]～[113]。各研究機関における会話音声認識システムの研究の概要を表1.4に示す。

表 1.4　会話音声認識の研究

(a) 外国

研究機関		CMU	CMU	CMU	CMU
システム名		Hearsay I	Dragon	Harpy	Hearsay II
対象 語彙数 発声環境 平均分岐数		チェス 24～194 端末室 —	チェス 24～194 端末室 —	AIに関する情報検索 1011 端末室 33.4	AIに関する情報検索 1011 端末室 33.4
文章例		Pawn to queen four.	Castle on king side.	What papers on grammatical inference are there ?	Which abstracts refer to theory of computation ?
音響処理	音響分析 音響処理結果 話者適応化	BPF分析（5ch） 零交差波 音韻系列 なし	BPF分析（5 ch） 零交差波 フレーム毎の音韻のスコア なし	LPC分析（14次） Itakura 尺度 音韻マトリクス＋スコア 有（20文章）	LPC分析 Itakura 尺度 音韻マトリクス＋スコア 有（60文章）
言語処理	単語認識 構文解析	記号のマッチング top-down	ネットワーク表現	ネットワーク表現，DP ネットワーク表現，beam search	グラフ表現，DP 書換え規則，island driven
性能	発声者数 文章総数 意味理解率 処理時間	4名 102文章 31% 実時間の9～44倍	1名 102文章 49% 実時間の48～174倍	5名 184文章 95% 実時間の80倍	1名 22文章 91% 実時間の240倍
稼動年月		1973	1975	1976	1976
備考		SUSの源	Harpyの第1次システム	ARPAのSUSでの成功システム	ブラックボードモデル

表 1.4　会話音声認識の研究（続き）

(a)　外国

研究機関		BBN	BBN	MIT	SRI
システム名		SPEECHLIS	HWIM	CASPERS	—
対象 語彙数 発声環境 平均分岐数		月の岩石に関する情報検索 250 無音室 —	旅行の予算管理 1097 端末室 196	— 236 端末室 9.15	水道栓の故障修理 54 無音室 —
文章例		Have any people done chemical analysis on this rock ?	What is the round trip fare to San Diego ?	Set the frequency range to 100Hz through 1000Hz.	Put one washer in the faucet.
音響処理	音響分析 音響処理結果 話者適応化	SLPC分析 セグメントラティス な し	SLPC分析（13次） Itakura 尺度 セグメントラティス＋スコア 声道長補正	LPC分析，ホルマント 音韻系列 な し	LPC分析，BPF分析 音響パラメータのデータベース な し
言語処理	単語認識 構文解析	ネットワーク表現， island driven	グラフ表現，tree search ネットワーク表現 best-first	グラフ表現 BN記法，best-first	word function BN記法，best-first
性能	発声者数 文章総数 意味理解率 処理時間	（HWIMに移行）	3名 124文章 44% 実時間の1350倍	6名 275文章 49% 実時間の15～20倍	3名 71文章 62% —
稼動年月		1975	1976	1974	1974
備考			SPEECHLISの改良		

表 1.4　会話音声認識の研究（続き）

(a) 外国

研究機関		SDC	IBM	IBM	BTL
システム名		VCDMS	—	CSAP 200	—
対象 語彙数 発声環境 平均分岐数		戦艦の情報検索 約1000 無音室 105	New Raleigh 言語 250 無音室 7.32	レーザーパテント 1000 無音室 24.1（Perplexity）	列車の座席予約 112 計算機室 20.0
文章例		Who is captain of the Wale ?	Some workers make the radio in use.	Suitable mounting means are well known in the art and are therefore not shown.	1日の，新大阪から，博多までの，6時22分の，ひかり19号の，グリーン券を，9枚，お願いします。
音響処理	音響分析 音響処理結果 話者適応化	LPC分析（24次），ホルマント，ピッチ Aマトリクス な し	FFT 音韻系列＋スコア な し	FFT（80次），ピッチ 音韻系列＋スコア 有（1時間音声）	ECLの音響処理
言語処理	単語認識 構文解析	グラフ表現 SRIを利用	グラフ表現，DP ネットワーク表現	グラフ表現，DP ネットワーク表現，best-first	ネットワークモデル DP
性能	発声者数 文章総数 意味理解率 処理時間	1名 54文章 24% （92MIPS）	1名 363文章 81% 実時間の300倍	1名 50文章 （文章認識率22%） 実時間の300倍	4名 80文章 79.4% 実時間の1000倍
稼動年月		1976	1976	1979	1980
備考		SRIと協同研究			ECLと共同研究

表 1.4　会話音声認識の研究（続き）

(a) 外国

研究機関		Instituto di Electrotecnica（伊）	CNET（仏）		
システム名		—	KEAL		
対象		交通機関の予約および案内	電話番号案内		
語彙数		約200	60		
発声環境		—	—		
平均分岐数		—	—		
文章例		—	Je voudrais le numero de telephone de MARCEL DUPONT.		
音響処理	音響分析	LPC分析，FFT，ホルマント，ピッチ	BPF分析（14次）		
	音響処理結果	音韻系列	音韻ラティス		
	話者適応化	なし	なし		
言語処理	単語認識		word spotting，DP		
	構文解析	BN記法，top-down	BN記法，top-down		
性能	発声者数	—	3名		
	文章総数	—	—		
	意味理解率	—	77%		
	処理時間	—	—		
稼動年月		1975	1977		
備考					

表 1.4　会話音声認識の研究（続き）

(b)　国内

研究機関		京都大学	京都大学	京都工芸繊維大学	山梨大学
システム名		LITHAN	—	SPOKEN—BASIC	—
対象 語彙数 発声環境 平均分岐数		計算機ネットワーク 101 無音室 5.0	デパートにおける会場案内 148 計算機室 7.8	BASIC プログラム 50 無音室 15.0	統計文を主とした日本語 99 無音室 —
文章例		計算機中央でいくつジョブは走っているか。	婦人靴売り場はどこですか。	100 DIM Y(10) CR	英語の得点の分散を求めよ。
音響処理	音響分析 音響処理結果 話者適応化	BPF分析（20ch） 音韻系列 なし	BPF分析（20ch） 音韻系列 なし	BPF分析（20ch） 音韻系列 なし	LPC分析(12次)，零交差波ホルマント 音韻系列 有
言語処理	単語認識 構文解析	音韻系列表現，DP BN記法，beam search	音韻系列表現，DP BN記法，beam search	グラフ表現，tree search ネットワーク表現，best-first	音韻系列表現，DP ネットワーク表現，depth-first
性能	発声者数 文章総数 意味理解率 処理時間	10名 200文章 64% —	1名 100文章 65% 実時間の数十倍	4名 242文章 85.5% —	4名 36文章 77.8% —
稼動年月		1975	1980	1975	1977
備考			オンラインで動作する質問回答システム		

表 1.4　会話音声認識の研究（続き）

(b) 国内

研究機関		ECL	ECL	ECL	
システム名		Voice Q-A System 0 （第1次システム）	Voice Q-A System I （第2次システム）	Voice Q-A System II	
対象 語彙数 発声環境 平均分岐数		列車の座席予約 65 無音室 —	列車の座席予約 112 計算機室 20.0	同左	
文章例		東京から，新大阪まで，8時30分発の，ひかり59号を，お願いします。	1日の，新大阪から，博多までの，6時22分の，ひかり19号の，グリーンを，9枚，予約します。	同左	
音響処理	音響分析 音響処理結果 話者適応化	相関分析（10次） 音韻ラティス な し	LPC分析（10次） Itakura 尺度 音韻ラティス＋継続時間 有（母音学習）	LPC分析 COSH尺度，WLR尺度 有（単独母音から推定）	
言語処理	単語認識 構文解析	音韻系列表現，tree search リスト表現，depth-first	高速tree search リスト表現，depth-first	同左	
性能	発声者数 文章総数 意味理解率 処理時間	1名 42文章（215文節） 87% 実時間の80倍	8名 320文章（1984文節） 86% 実時間の5倍	9名 180文章（1116文節） 96.9% 実時間の9倍	
稼動年月		1975	1976	1980	
備考			オンライン質問回答システム		

1.3 日本語音声認識の問題点と研究方針

本論文は，日本語音声を認識対象とした認識システムの作成を目標として著者が行ってきた研究の報告である。

日本語音声を対象とした認識システムを構成するにあたっては少くとも次の点について検討を行う必要がある。

(1) 認識対象

(2) 音声の個人差

(3) 入力音声の音響分析

(4) 標準パターンの表現単位

(5) 単語音声認識アルゴリズム

(6) 連続単語音声認識アルゴリズム

(7) 会話音声認識アルゴリズム

(8) アルゴリズムの高速化，装置化

(1)に関しては，音声認識の最終目標は会話音声の認識であるが，最初から会話音声認識をねらうのは，音声認識の困難さからして危険が大きい。また，実用的な観点からすると，単語音声のような比較的簡単な対象を取り扱うことも十分意味のあることである。そこで本論文では認識対象を，単語音声，連続単語音声，会話音声の3種類に設定し，これらを順次取り扱うこととする。これらのうち単語音声は，数字，地名などの単語を区切って発声したものをさしており，音声認識の対象として最初に取り上げるべきものである。また単語音声認識の段階でも，その実用面での応用範囲は非常に広範囲にわたっている。次の連続単語音声とは，数字などの単語を区切らずに連続して発声した音声である。これは単語が複数個連続したものを取り扱うという意味で，単語音声から一歩進んだものであり，単語音声から会話音声へ進む中間段階として適当な単位である。また，単語を区切らずに連続して発声することが出来れば，発声者の負担が軽減され，かつ情報発生速度も速くなるから，実用的観点からも大きな意味を持っている。最後の会話音声認識は音声認識の最終目標であり，会話音声を取り扱ってどの程度の性能が得られるかは，認識アルゴリズムの正当性の検証の点からも大きな意味をもつ。これらの単語音声，連続単語音声，会話音声を一貫した手法で取り扱うことが望ましいことはもちろんであるが，それぞれの対象に適した認識手法を取ることが，高い認識性能を得る条件であることも事実である。そこで本論文では後で述べるように2種類の認識手法を提案して検討を

行う。

(2)に関しては，認識システムを不特定多数の発声者に対して良好に動作させようとすると，音声の特徴量の個人によるバラツキが大きな問題となる。音声の個人性に関する規則が明らかになれば，音声波から個人差に関係した特徴量を取り去ってやればよいが，これは極めて困難な問題であって，音声の個人性に関する研究はまだ暗中模索の段階である。したがって本論文では音声の個人性の問題には直接ふれず，標準パターンを発声者ごとに用意するという方法でこの問題に対処することにする。

(3)の項目は，音声認識の基本的な部分であり認識性能を左右する大きな要素である。これに関しては，音声のスペクトル包絡を時間領域で抽出する最近の新しい分析法が，アルゴリズムの明解さ，分析精度，ディジタル処理との適合性の点から優れていると考えられる。そのような分析法として，最尤スペクトル分析，線形予測分析（ＬＰＣ分析），PARCOR分析などがある。これらは数学的には等価であるが，特徴パラメータの表現形式が異なっている。音声認識のための分析法としては，スペクトル間の距離（又は類似度）が簡単に計算されて，その意味がわかりやすいことが望ましい。最尤スペクトル分析では，スペクトル間の類似度が簡単な計算で求まり，しかもその値が統計的推定論における対数尤度に対応しているので，望ましい条件を満している。したがって，本論文では最尤スペクトル分析法を用いることとする。

(4)に関しては標準パターンを単語単位で表現する方法と，音韻，音節のような単語より小さい単位で表現する方法がある。単語単位で表現する方法については，音声の基本的な単位が単語とは考えられないこと，認識対象の単語が多くなると標準パターンの登録が困難になること，会話音声認識への拡張性に乏しいことなどの理由で採用しないこととする。単語より小さい単位で標準パターンを表現する方法として本論文では２つの方法を用いる。１つは，母音－子音－母音より構成される音節 ―― 以下本論文中ではこれをＶＣＶ音節と呼ぶ ―― を標準パターンを表現する単位とする方法であり，他の１つは音韻を表現単位とする方法である。まず，ＶＣＶ音節を単位とする方法では，認識の際，セグメンテーション→認識というオーソドックスな手法をとることにする。通常良く使われる音韻を単位にとる方法に比較して，ＶＣＶ音節を単位にとる方法の特徴をあげると以下のようになる。

(a)　ＶＣＶ音節を単位としてセグメンテーションを行うには，母音部分を検出すればよい。音声中の母音部分の検出は比較的容易である。これに対し，音韻を単位とした方法では，母音部分，子音部分を正しく検出する必要があるが，子音は一般に定常部分が存在しないので検出は困難である。

(b) 音韻を単位とする方法で子音の認識を行う場合には，時間軸上のある一点におけるスペクトルパターンのみから認識を行うことが多い。しかしながら，子音の特徴はその前後の母音との間のスペクトルパターンの特有な推移にあるから，子音を前後の母音と組み合わせたVCV音節を音声単位にとることにより，子音の認識率が向上すると予測される。

(c) 日本語はＶＣＶ……という音韻構成になっているから，ＶＣＶ音節を音声を構成している単位であると考えても矛盾はない。

これに対し，音韻を標準パターンの単位とする場合には，セグメンテーション→認識という手順をふむ方法をとると，上に述べたようにＶＣＶ音節を単位とした方法に比較して不利なことは明らかである。そこで，この場合は認識対象を単語音声，もしくは連続単語音声に限ることとし，セグメンテーションは行わず，入力音声と各音韻標準パターンとの類似度を求め，次に音韻系列として表現してある単語辞書とのマッチングにより認識するという手法をとる。この方法は標準パターンは音韻単位で表現し，認識は単語単位で行う方法であり，標準パターンを単語単位で表現する方法に比較して，次のような利点を特っている。

(a) 認識語彙の増加，変更の際は単語辞書の内容のみ変更すればよいため，語彙の増加，変更に適応しやすい。

(b) 発声者が交替する際は，音韻標準パターンの作り直しが必要であるが，音韻の種類は数十個程度と比較的少ないので，標準パターンを作り直すための処理を簡略化できる可能性がある。

(c) 単語を標準パターンの単位とする方法に比較して，標準パターン，単語辞書を蓄えるための記憶容量が少なくてすむ。

以上のような理由で，本論文では標準パターンの単位として，ＶＣＶ音節，音韻の２種類を用いることとする。そして，ＶＣＶ音節を単位とする方法では，単語音声，連続単語音声，会話音声の認識を行う。また，音韻を単位とする方法では，単語音声，連続単語音声の認識を行う。

(5)に関しては，最も重要なのは発声速度の変動に対処するために，時間軸の正規化法について考慮することである。というのは，先に述べたように同じ発声者が同じ単語を発声した場合でも，発声するたびに得られる音声パターンが異なり，その主たる要因が発声速度の変動によるものだからである。この変動は，時間軸の局所的な伸縮としてあらわれるから，発声速度の変動に対処するには時間軸の非線形正規化を行ったマッチングを行う必要がある。これについてはＤＰを用いるマッチング法の有効さが示されているので，これを用いることとし，ＶＣＶ

音節，音韻を標準パターンに用いるそれぞれの場合に適したＤＰの方法について検討する。また，ＶＣＶ音節を単位とした方法では，セグメンテーションが必要になるが，セグメンテーションの際生じる誤りは認識の段階で回復することが困難である。そこで，セグメンテーションの際に，あいまいさを持たせることを許すことにして，誤りが生じることを防ぐ方法をとる。

(6)に関しては，その基本となる単語音声認識アルゴリズムと，それを拡張して連続単語音声を認識するアルゴリズムが重要である。まず，単語音声認識アルゴリズムは，(5)で述べたことがそのままあてはまるから，(5)と同じ方針をとる。次に単語音声認識をベースにして連続単語音声認識を行う際，通常の方法は，連続単語音声を単語単位にセグメンテーションし，しかる後，単語音声認識を行うという方法である。しかしながら，連続単語音声中の単語境界は物理的に明確には現われないため，セグメンテーションは極めて困難である。そこで本論文では，すべての可能な単語系列と入力音声とのマッチングを行い，最も良くマッチングされた単語系列を認識結果とする方法をとる。また，この際，膨大な計算量の削減をはかるためＤＰを導入する。

(7)に関しては，何を目標とするかが最も重要な問題である。認識対象を特に限定しないいわゆる音声タイプライタの実現は現時点では困難である。そこで，音声理解の考え方を取り入れ，タスクを特定の領域に限定し，その中で認識システム全体としての性能向上をめざすことにする。もちろんそのためには，音響処理の性能が高いことが必要であり，セグメンテーション，音韻認識の性能向上をはかる。さらに重要なことは，音響処理結果の表現形式である。言語処理の際，扱いやすくしかも十分な情報を持っている必要がある。従来の方法では，音響処理結果を音韻系列で表現するのが通常であったが，これでは，セグメンテーション，認識の際に生じる誤りをそのまま言語処理に送ることになり，言語処理部での誤り訂正が困難になる。これに対し本論文では，セグメンテーション，認識の結果にあいまいさを持たせて音響処理結果を表現することにし，言語処理部の負担を軽減し，会話音声認識システム全体の性能向上を計る。

(8)に関しては，音声認識研究の目標が，人間と機械の間の音声による情報伝達手段の実現にある以上，アルゴリズムの有効性を示すには，実際に認識システム，認識装置を作成して，それらが良好に動作することを確かめる必要がある。そこで本論文では，単語音声認識，連続単語音声認識に関しては，それぞれ装置を試作して性能評価を行う。また会話音声認識に関しては，オンラインで動作する認識システムを構成して性能評価を行う。

次に，本論文の構成について述べる。まず第２章では，最尤スペクトル分析法を用いた音声分析系の構成について述べる。第３章では，ＶＣＶ音節を単位とした音声認識の基礎として，

単独に発声されたＶＣＶ音節の認識について述べる。第４章ではＶＣＶ音節を単位として単語音声を認識する方法と認識実験による評価について述べる。第５章では，ＶＣＶ音節を単位とした連続単語音声の認識法を提案し，認識実験により評価を行う。第６章ではＶＣＶ音節の認識法の改良について述べ，それを連続単語音声認識に応用して効果を明らかにする。第７章では対象を会話音声に拡張する。まず会話の対象について述べ，音響処理部，言語処理部の構成について述べる。さらに，会話音声認識システムの性能評価実験について述べる。第８章では第７章で述べた会話音声認識システムに改良を加えた第２次システムの構成および性能評価実験について述べる。第９章，第10章では標準パターンの表現単位として音韻を用いた場合の認識方法について述べる。まず第９章では，音韻を標準パターンの単位にとった場合の単語音声認識法を提案し，認識実験による評価について述べる。第10章では，対象を連続単語音声に拡張し，３種類の連続単語音声認識法を提案し，認識実験によりそれらの比較評価を行う。第11章では，本論文で提案したアルゴリズムの高速化，装置化を試みる。まず，試作した単語音声認識装置，連続単語音声認識装置について述べる。さらに，オンラインで，かつ実時間の数倍以内の処理時間で動作する会話音声認識システムについて述べる。

第2章　音響分析

2.1　はしがき

本章では，本論文で用いている音響分析法について述べる。まず，本論文で用いている音響分析法である最尤スペクトル分析法について述べる。次に，入力パターンと標準パターンの距離を求める際に用いる距離尺度について述べる。最後に，最尤スペクトル分析法に基づく入力音声の具体的な音響分析法について述べる。

2.2　最尤スペクトル分析法[16]

最尤スペクトル分析法とは，音声信号を全極モデルで表現して，そのモデルのパラメータを統計的推定論における最尤法で推定する分析法のことである。

音声波形の現時点の標本値 $y(n)$と，これに隣接する過去の p 個の標本値との間に次の線形結合関係が成立すると仮定する。

$$y(n)+\alpha_1 y(n-1)+\alpha_2 y(n-2)+\cdots\cdots+\alpha_p y(n-p)=\sigma\varepsilon(n) \qquad (2.1)$$

ただし，$\sigma\varepsilon(n)$ は駆動音源入力信号であり，σ はそのRMS値，$\varepsilon(n)$は単位RMS値をもち，平坦なスペクトルを持つ信号であると仮定する。α_i は過去の信号より現在の信号を

$$\tilde{y}(n)=-\sum_{i=1}^{p}\alpha_i y(n-i) \qquad (2.2)$$

として予測する係数であり，線形予測係数と呼ばれる。また，

$$y(n)-\tilde{y}(n)=\sigma\varepsilon(n) \qquad (2.3)$$

であるから，$\sigma\varepsilon(n)$は線形予測による予測誤差（もしくは予測残差）である。

式（2.1）の差分方程式を満す $\varepsilon(n)$から $y(n)$への伝達関数は

$$H(Z)=\frac{\sigma}{1+\sum_{i=1}^{p}\alpha_i Z^i} \qquad (2.4)$$

で与えられ，かつ，そのパワースペクトル密度$T(w)$は

$$T(w)=\frac{1}{2\pi}\mid H(Z)\mid^2$$

$$=\frac{1}{2\pi}\frac{\sigma^2}{\mid 1+\sum_{i=1}^{p}\alpha_i Z^i\mid^2} \qquad (2.5)$$

となる。ただし，$Z=e^{-jw}$（$-\pi\leqq w\leqq\pi$）である。すなわち，$T(w)$は，零点を持たない有理スペクトル（全極形有理スペクトル）で表現される。最尤法では，$\varepsilon(n)$を定常的白色ガウス雑音であると仮定し，この仮定に基づき，短時間区間の音声波形$Y=(y(1), y(2), \cdots\cdots, y(N))$から，パラメータ（$\sigma^2, \alpha_1, \alpha_2, \cdots\cdots, \alpha_p$）の最適な推定を行う。推定は，パラメータが与えられたときの標本値が実現する尤度を最大化することによって行われる。

尤度を最大化する際，次式で定義される対数尤度を用いる。

$$l(Y;\sigma^2,\{\alpha_i\})=-\frac{N}{2}\left\{\log 2\pi\sigma^2+\frac{1}{\sigma^2}\sum_{i=0}^{p}\sum_{j=0}^{p}\alpha_i\alpha_j u_{i-j}\right\}$$

$$=-\frac{N}{2}\left[2\log 2\pi+\frac{1}{2\pi}\left\{\log T(w)+\frac{S(w)}{T(w)}\right\}dw\right] \qquad (2.6)$$

ただしu_iは

$$u_i=\frac{1}{N}\sum_{t=1}^{N-i}y_t y_{t+i} \qquad (2.7)$$

で定義される音声波形Yの短時間自己相関関数であり，$S(w)$はその短時間スペクトルである。また，$T(w)$は式（2.5）で定義される全極形のパワースペクトル密度である。

$l(Y;\sigma^2,\{\alpha_i\})$を最大化する$\sigma^2$と$\{\alpha_i\}$を求める。まず，$\sigma^2$について最大化するには，$l(Y;\sigma^2,\{\alpha_i\})$の$\sigma^2$に関する微係数を0とおくことにより

$$\sigma^2=\sum_{i=0}^{p}\sum_{j=0}^{p}\alpha_i\alpha_j u_{i-j} \qquad (\alpha_0=1) \qquad (2.8)$$

のとき最大値をとり，その値は次のようになる。

$$\max_{\sigma^2} l\,(Y\,;\,\sigma^2,\,\{\,\alpha_i\,\}) = -\frac{N}{2}\left\{1+\log 2\pi+\right.$$

$$\left.\log\left(\sum_{i=0}^{p}\sum_{j=0}^{p}\alpha_i\,\alpha_j\,u_{i-j}\right)\right\} \qquad (2.9)$$

さらに，α_iについては，σ^2をα_iについて最小化することに等しいから，式（2.8）のα_iに関する微係数を0とおくと，連立一次方程式

$$\sum_{i=1}^{p}\alpha_i\,u_{i-j}=0 \qquad (j=1,\,2,\,\cdots\cdots,\,p) \qquad (2.10)$$

が得られる。またこのとき最大値をl maxとすると，これは次式で与えられる。

$$l\max = -\frac{N}{2}\left\{1+2\pi+\log\left(\sum_{i=1}^{p}\alpha_i\,u_i\right)\right\} \qquad (2.11)$$

一方，式（2.5）の$T(w)$は次式のように表わすこともできる。

$$T(w)=\frac{\sigma^2}{2\pi}\,\frac{1}{\sum_{k=-p}^{p}A_k\,e^{jkw}}$$

$$=\frac{\sigma^2}{2\pi}\,\frac{1}{A_0+2A_1\cos w+\cdots\cdots+2A_p\cos pw} \qquad (2.12)$$

ただし，

$$A_i=\sum_{j=0}^{p-i}\alpha_j\,\alpha_{j+i} \qquad (2.13)$$

A_iは線形予測係数の自己相関関数であり，一般に最尤スペクトルパラメータと呼ばれる。

以上のことから，音声の標本値Yから式（2.7）によって得られる自己相関関数，式（2.13）によって得られる最尤スペクトルパラメータは，いずれも音声の特徴を表わすパラメータであることがわかる。

最尤スペクトル推定法の物理的意味は次のように説明される。式（2.6）において，$l\,(Y\,;\,\sigma^2,\,\{\,\alpha_i\,\})$の最大化を$T(w)$について行うと，

$$l\max = l\,(Y\,;\,\sigma^2,\,\{\,\alpha_i\,\})\,|_{\,T(w)=S(w)}$$

$$=-\frac{N}{2}\left\{2\log 2+\frac{1}{2\pi}\int(\log S(w)+1)\;dw\right\} \qquad (2.14)$$

したがって，$l\,\max - l(Y;\sigma^2,\{\alpha_i\})$ を求めると，

$$l\,\max - l(Y;\sigma^2,\{\alpha_i\}) =$$

$$\frac{N}{8\pi}\int_{-\pi}^{\pi} 2\left\{\log\frac{T(w)}{S(w)}+\frac{S(w)}{T(w)}-1\right\}dw \qquad (2.15)$$

となり，右辺の

$$E=\int_{-\pi}^{\pi} 2\left\{\log\frac{T(w)}{S(w)}+\frac{S(w)}{T(w)}-1\right\}dw \qquad (2.16)$$

は，短時間スペクトルを全極形の有理スペクトルでおきかえたときの，誤差の評価尺度になる。

$$D(w)=\log T(w)-\log S(w) \qquad (2.17)$$

とおくと，式（2.16）は

$$E=2\int_{-\pi}^{\pi}\left\{D(w)+e^{-D(w)}-1\quad dw\right\} \qquad (2.18)$$

となる。式（2.18）の被積分項を$D(w)=0$のまわりでテーラー展開すると

$$D(w)+e^{-D(w)}-1=\sum_{i=2}^{\infty}\frac{1}{i!}(-D(w))^i$$

$$=\frac{1}{2}D^2(w)-\frac{1}{6}D^3(w)+\frac{1}{24}D^4(w)-\cdots\cdots \qquad (2.19)$$

となる。したがって，Eは$|D(w)|\ll 1$のとき対数スペクトルの2乗誤差尺度に近似的に等しい。一方，$|D(w)|>1$の場合には$D(w)$が正なら$D(w)$に比例する誤差尺度となり，$D(w)$が負の場合には指数関数$e^{|D(w)|}$の重みを持つことがわかる。

2.3 距離尺度（類似度尺度）

本節では，標準パターンと入力パターンの間の距離を定義する距離尺度について述べる。前節の式（2.15）は，短時間スペクトル$S(w)$を全極形の有理スペクトル$T(w)$でおきかえたときの誤差の距離尺度であった。この関係式を入力パターンと標準パターン間の距離尺度を定義するのに用いる。いま，標準パターンの最尤スペクトル分析法により抽出したスペクトル包絡を$T(w)$

とする。また$T(w)$に対応する，相関関数，線形予測パラメータ，最尤スペクトルパラメータ，残差パワーを，u_i，α_i，A_i，σ_T^2とする。同様に入力音声のスペクトル包絡を$R(w)$とし，$R(w)$に対応する相関関数，線形予測パラメータ，最尤スペクトルパラメータ，残差パワーを，v_i，β_i，B_i，σ_R^2とする。これらのパラメータの関係を表2.1にまとめておく。

$T(w)$，$R(w)$のスペクトル包絡は次式で表わされる。

表2.1　記号の説明

	スペクトル包絡	自己相関関数	線形予測パラメータ	最尤スペクトルパラメータ	残差パワー
標準パターン	$T(w)$	u_i	α_i	A_i	σ_T^2
入力音声	$R(w)$	v_i	β_i	B_i	σ_R^2

$$T(w)=\frac{1}{2\pi}\frac{\sigma_T^2}{|1+\alpha_1 Z+\cdots\cdots+\alpha_p Z^p|^2}$$

$$=\frac{1}{2\pi}\frac{\sigma_T^2}{\sum_{k=-p}^{p} A_k e^{jkw}} \qquad (2.20)$$

$$R(w)=\frac{1}{2\pi}\frac{\sigma_R^2}{|1+\beta_1 Z+\cdots\cdots\beta_p Z^p|^2}$$

$$=\frac{1}{2\pi}\frac{\sigma_R^2}{\sum_{k=-p}^{p} B_k e^{jkw}} \qquad (2.21)$$

このとき$T(w)$と$R(w)$の間の距離$d(T, R)$を次式で定義する。

$$d(T,R)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log\frac{T(w)}{R(w)}+\frac{R(w)}{T(w)}-1\right\}dw \qquad (2.22)$$

$T(w)$のすべての極の絶対値$|Z_i|$が1より大きいことにより$\log T(w)$を$[-\pi, \pi]$で定積分すれば次の関係式が得られる。

$$\int_{-\pi}^{\pi} \log T(w)\,dw = 2\pi \log \frac{\sigma_T^2}{2\pi} \qquad (2.23)$$

同様にして次の関係式も得られる。

$$\int_{-\pi}^{\pi} \log R(w)\,dw = 2\pi \log \frac{\sigma_R^2}{2\pi} \qquad (2.24)$$

さらに関係式

$$R(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} v_k\, e^{jkw} \qquad (2.25)$$

（ウィーナ・ヒンチンの定理）を用いると，

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{R(w)}{T(w)}dw = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\sigma_T^2}\left(\sum_{k=-p}^{p} A_k\, e^{jkw} \sum_{k=-\infty}^{\infty} v_k\, e^{jkw}\right) dw$$

$$= \frac{1}{\sigma_T^2} \sum_{k=-p}^{p} A_k\, v_k \qquad (2.26)$$

式（2.23）～式（2.26）を式（2.22）に代入すると$d(T, R)$は次式で与えられる。

$$d(T, R) = \log \frac{\sigma_T^2}{\sigma_R^2} + \frac{1}{\sigma_T^2} \sum_{k=-p}^{p} A_k\, v_k - 1 \qquad (2.27)$$

次に$d(T, R)$の値が最小になるようにσ_T，σ_Rの値を設定する。そのため，$d(T, R)$をσ_Tで偏微分して0とおく。[16][19]

$$\frac{\partial l(T, R)}{\partial(\sigma_T^2)} = \frac{1}{\sigma_T^2} - \frac{1}{(\sigma_T^2)^2} \sum_{k=-p}^{p} A_k\, v_k = 0 \qquad (2.28)$$

$$\sigma_T^2 = \sum_{k=-p}^{p} A_k\, v_k \qquad (2.29)$$

したがって，$d(T, R)$は

$$d(T, R) = \log \frac{\sigma_T^2}{\sigma_R^2}$$

$$= \log \sum_{k=-p}^{p} A_k v_k - \log \sum_{k=-p}^{p} B_k v_k \qquad (2.30)$$

となる。式（2.30）の右辺第2項は入力パターンのみに関係した項である。したがって入力パターンと各標準パターンの絶対的な距離ではなく，相対的な距離のみを問題にするときは，距離としては式（2.30）の代りに

$$d(T, R) = \log \sum_{k=-p}^{p} A_k v_k \qquad (2.31)$$

を用いることもできる。さらに式（2.31）を書きかえると

$$d(T, R) = \log \left(A_0 v_0 + 2 \sum_{k=1}^{p} A_k v_k \right) \qquad (2.32)$$

であるから，$2A_i$（$i \geqq 1$）を改めてA_iとおくことにより，式（2.32）は

$$d(T, R) = \log \sum_{k=0}^{p} A_k v_k \qquad (2.33)$$

と簡単化することができる。また式（2.33）の符号を逆にした

$$l(T, R) = -\log \sum_{k=0}^{p} A_k v_k \qquad (2.34)$$

は，標準パターンと入力パターンの類似度と考えることができる。また，式（2.30）で自己相関関数を自己相関係数

$$\rho_i = \frac{v_i}{v_0} \qquad (2.35)$$

でおきかえても良いことは明らかであるから，式（2.30）～式（3.34）は自己相関関数の代りに自己相関係数を用いても良い。本論文中では主として式（2.34）を用いる。

2.4 音響分析系

ここでは，最尤スペクトル分析法に基づく音響分析系について述べる。音響分析系のブロック図を図2.1に示す。

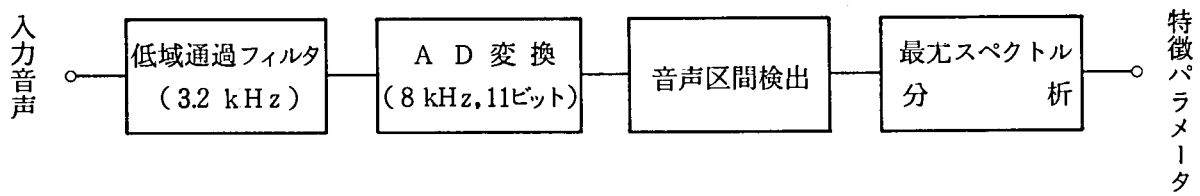


図 2.1　音響分析系の構成

2.4.1 標本化

入力音声を，まず，3.2 kHz のしゃ断周波数を持つ低域フィルタに通す。この周波数帯域は，電話系の周波数帯域にほぼ等しい。次に，標本化周波数8 kHzのA D変換器で符号＋11ビットのディジタル音声に変換する。

2.4.2 音声区間検出

次に，背景雑音の中から音声区間を抽出する。ディジタル音声を15 msec ごとに区分し，各区分（フレーム）ごとの音声パワーを求める。求められたパワーが閾値をこえたかどうかで音声区間検出を行う。すなわち，音声区間の始端は，音声パワーが始めて閾値をこえたフレームとする。また，終端は，閾値以下の音声パワーがT_E msec（T_Eは定数）以上続いたとき，音声パワーが閾値以下になる直前のフレームとする。（図2.2）

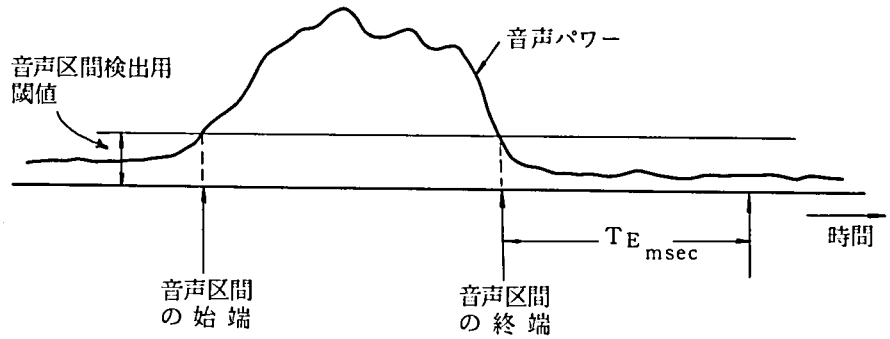


図 2.2　音声区間の検出

音声区間の途中に無声子音などがあるとその直前に無音区間が生じる。ここでは，音声区間の途中で音声パワーが閾値以下になった区間を無音区間と呼ぶことにする。(図 2.3)

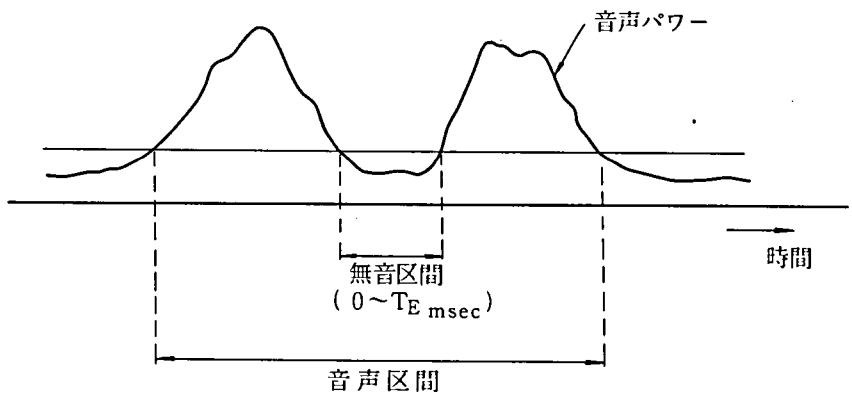


図 2.3　音声区間中の無音区間

2.4.3　最尤スペクトル分析

次に，上で検出された音声区間における音声波の特徴量を抽出するために，最尤スペクトル分析を行う。分析の条件は以下の通りである。

分析周期：　15 msec

分析フレーム長：　15 msec

分析窓のタイプ：　矩形窓

分析の次数：　p (= 10)

音響分析部の処理結果は，音声信号の自己相関関数，および，最尤スペクトルパラメータの時系列で表わされる。第i番目のフレームの自己相関関数を

$$\boldsymbol{v}_i = (v_{i0}, v_{i1}, \cdots\cdots, v_{ip}) \qquad (2.36)$$

最尤スペストルパラメータを

$$\boldsymbol{B}_i = (B_{i0}, B_{i1}, \cdots\cdots, B_{ip}) \qquad (2.37)$$

とすると，入力音声は

$$(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\cdots, \boldsymbol{v}_N) \qquad (2.38)$$

$$(\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots\cdots, \boldsymbol{B}_N)$$

と表現される。ただしNは音声区間のフレーム数である。なお，距離尺度（類似度尺度）として式（2.33）又は式（2.34）を用いる場合には自己相関関数のみで良く，最尤スペクトルパラメータは必要ない。

2.5 あとがき

本章では，本論文で用いる音響分析法について述べた。音響分析法としては，最尤スペクトル分析法を用いた。この方法によれば，入力音声の特徴パラメータとして，音声波形の自己相関関数，最尤スペクトルパラメータ等を用いることができることを示した。また，最尤スペクトル分析法に基づいて，分析された入力パターンと標準パターンの間の距離を定義する距離尺度について述べた。その結果，標準パターンは最尤スペクトルパラメータの形で蓄えておけばよいことを示した。また，入力パターンについては，自己相関関数と最尤スペクトルパラメータで表現すれば良いことを示した。特に，相対的な距離値のみを問題とする場合には，自己相関関数のみで良く，これと，標準パターンの最尤スペクトルパラメータとの積和計算で容易に距離計算が行えることを示した。

次に，最尤スペクトル分析法を用いた音響析系について述べた。系の構成，各部の処理，具体的な分析条件を示し，本論文で用いられている音響分析系を明確にした。

第3章　母音－子音－母音型音節(VCV音節)の認識

3.1　はしがき

第1章で述べたように，母音，子音，母音より構成される音節（ＶＣＶ音節）は，日本語を構成している基本的な単位と考えて良いと同時に次のような理由で日本語音声を認識する際の単位として適当である。

(1)　セグメンテーションが行いやすい。

(2)　子音の認識が行いやすい。

さらに，ＶＣＶ音節を単位とした日本語文章の合成が試みられており，[114][115]その結果良好な音声が合成されていることも，合成という立場からこれらのことを裏付けているといえる。

本章では，ＶＣＶ音節を単位とした日本語音声の認識の手はじめとして，ＶＣＶ音節の認識の問題を取り扱うことにする。日本語に現われるVCV音節は約900種類あるが，ここでは，それらのうち重要なもの300種類を認識対象とする。また，問題を簡単化するため，ここでは，単独に発声されたＶＣＶ音節を取り扱う。連続音声中に現われるＶＣＶ音節の認識については第6章で論じる。

ＶＣＶ音節の認識を行う際には

(1)　標準パターンの作成法

(2)　標準パターンを用いて入力音声を認識する方法

の2点が重要である。(1)については，いくつかの学習サンプルを，時間軸の正規化を行った後平均することにより標準パターンを作成する方法を提案し，他の音声単位との比較等を行う。(2)については，時間軸の正規化を行ったパターンマッチング法[33][34][35]を採用し，種々の制限をつけることにより認識率の向上をはかる。

3.2　認識系の構成

3.2.1　認識対象

認識対象は日本語の母音5種（/a/，/i/，/u/，/e/，/o/）および子音12種（/m/，/n/，

/b/, /d/, /g/, /r/, /z/, /p/, /t/, /k/, /s/, /h/) よりなる300種類のＶＣＶ音節である。これらのＶＣＶ音節のリストを表3.1に示す。

表3.1　認識対象となるＶＣＶ音節

m	ama（アマ）	ami（アミ）	amu（アム）	ame（アメ）	amo（アモ）
	ima（イマ）	imi（イミ）	imu（イム）	ime（イメ）	imo（イモ）
	uma（ウマ）	umi（ウミ）	umu（ウム）	ume（ウメ）	umo（ウモ）
	ema（エマ）	emi（エミ）	emu（エム）	eme（エメ）	emo（エモ）
	oma（オマ）	omi（オミ）	omu（オム）	ome（オメ）	omo（オモ）

n	ana（アナ）	ani（アニ）	anu（アヌ）	ane（アネ）	ano（アノ）
	ina（イナ）	ini（イニ）	inu（イヌ）	ine（イネ）	ino（イノ）
	una（ウナ）	uni（ウニ）	unu（ウヌ）	une（ウネ）	uno（ウノ）
	ena（エナ）	eni（エニ）	enu（エヌ）	ene（エネ）	eno（エノ）
	ona（オナ）	oni（オニ）	onu（オヌ）	one（オネ）	ono（オノ）

b	aba（アバ）	abi（アビ）	abu（アブ）	abe（アベ）	abo（アボ）
	iba（イバ）	ibi（イビ）	ibu（イブ）	ibe（イベ）	ibo（イボ）
	uba（ウバ）	ubi（ウビ）	ubu（ウブ）	ube（ウベ）	ubo（ウボ）
	eba（エバ）	ebi（エビ）	ebu（エブ）	ebe（エベ）	ebo（エボ）
	oba（オバ）	obi（オビ）	obu（オブ）	obe（オベ）	obo（オボ）

d	ada（アダ）	adi（アディ）	adu（アドゥ）	ade（アデ）	ado（アド）
	ida（イダ）	idi（イディ）	idu（イドゥ）	ide（イデ）	ido（イド）
	uda（ウダ）	udi（ウディ）	udu（ウドゥ）	ude（ウデ）	udo（ウド）
	eda（エダ）	edi（エディ）	edu（エドゥ）	ede（エデ）	edo（エド）
	oda（オダ）	odi（オディ）	odu（オドゥ）	ode（オデ）	odo（オド）

g	aga（アガ）	agi（アギ）	agu（アグ）	age（アゲ）	ago（アゴ）
	iga（イガ）	igi（イギ）	igu（イグ）	ige（イゲ）	igo（イゴ）
	uga（ウガ）	ugi（ウギ）	ugu（ウグ）	uge（ウゲ）	ugo（ウゴ）
	ega（エガ）	egi（エギ）	egu（エグ）	ege（エゲ）	ego（エゴ）
	oga（オガ）	ogi（オギ）	ogu（オグ）	oge（オゲ）	ogo（オゴ）

	ara (アラ)	ari (アリ)	aru (アル)	are (アレ)	aro (アロ)
	ira (イラ)	iri (イリ)	iru (イル)	ire (イレ)	iro (イロ)
r	ura (ウラ)	uri (ウリ)	uru (ウル)	ure (ウレ)	uro (ウロ)
	era (エラ)	eri (エリ)	eru (エル)	ere (エレ)	ero (エロ)
	ora (オラ)	ori (オリ)	oru (オル)	ore (オレ)	oro (オロ)

	aza (アザ)	azi (アジ)	azu (アズ)	aze (アゼ)	azo (アゾ)
	iza (イザ)	izi (イジ)	izu (イズ)	ize (イゼ)	izo (イゾ)
z	uza (ウザ)	uzi (ウジ)	uzu (ウズ)	uze (ウゼ)	uzo (ウゾ)
	eza (エザ)	ezi (エジ)	ezu (エズ)	eze (エゼ)	ezo (エゾ)
	oza (オザ)	ozi (オジ)	ozu (オズ)	oze (オゼ)	ozo (オゾ)

	apa (アパ)	api (アピ)	apu (アプ)	ape (アペ)	apo (アポ)
	ipa (イパ)	ipi (イピ)	ipu (イプ)	ipe (イペ)	ipo (イポ)
p	upa (ウパ)	upi (ウピ)	upu (ウプ)	upe (ウペ)	upo (ウポ)
	epa (エパ)	epi (エピ)	epu (エプ)	epe (エペ)	epo (エポ)
	opa (オパ)	opi (オピ)	opu (オプ)	ope (オペ)	opo (オポ)

	ata (アタ)	at∫i (アチ)	atsu(アツ)	ate (アテ)	ato (アト)
	ita (イタ)	it∫i (イチ)	itsu(イツ)	ite (イテ)	ito (イト)
t	uta (ウタ)	ut∫i (ウチ)	utsu(ウツ)	ute (ウテ)	uto (ウト)
	eta (エタ)	et∫i (エチ)	etsu(エツ)	ete (エテ)	eto (エト)
	ota (オタ)	ot∫i (オチ)	otsu(オツ)	ote (オテ)	oto (オト)

	aka (アカ)	aki (アキ)	aku (アク)	ake (アケ)	ako (アコ)
	ika (イカ)	iki (イキ)	iku (イク)	ike (イケ)	iko (イコ)
k	uka (ウカ)	uki (ウキ)	uku (ウク)	uke (ウケ)	uko (ウコ)
	eka (エカ)	eki (エキ)	eku (エク)	eke (エケ)	eko (エコ)
	oka (オカ)	oki (オキ)	oku (オク)	oke (オケ)	oko (オコ)

	asa (アサ)	a∫i (アシ)	asu (アス)	ase (アセ)	aso (アソ)
	isa (イサ)	i∫i (イシ)	isu (イス)	ise (イセ)	iso (イソ)
s	usa (ウサ)	u∫i (ウシ)	usu (ウス)	use (ウセ)	uso (ウソ)
	esa (エサ)	e∫i (エシ)	esu (エス)	ese (エセ)	eso (エソ)
	osa (オサ)	o∫i (オシ)	osu (オス)	ose (オセ)	oso (オソ)

	aha (アハ)	ahi (アヒ)	ahu (アフ)	ahe (アヘ)	aho (アホ)
	iha (イハ)	ihi (イヒ)	ihu (イフ)	ihe (イヘ)	iho (イホ)
h	uha (ウハ)	uhi (ウヒ)	uhu (ウフ)	uhe (ウヘ)	uho (ウホ)
	eha (エハ)	ehi (エヒ)	ehu (エフ)	ehe (エヘ)	eho (エホ)
	oha (オハ)	ohi (オヒ)	ohu (オフ)	ohe (オヘ)	oho (オホ)

3.2.2 音声の表現

入力音声の分析は第2章で述べたのと同じ方法で行う。入力音声はまず3.2kHz低域通過フィルタを通ったあと，標準化周波数8kHzのAD変換器で11ビットのディジタル音声に変換される。次に，このディジタル音声を15msecごとに区分し，各フレームの音声パワーを計算する。あらかじめ定められた閾値以上になったフレームを音声区間の始端とする。閾値以下の音声パワーが225msec（15フレーム）以上続いた場合は，最初に閾値以下になったフレームを音声区間の終端とする。音声区間内で閾値以下の音声パワーが15msec以上，225msec以下続く部分は無音区間とする

以上のようにして切り出されたVCV音節の音声区間の周波数スペクトルを表わす特徴量として，各フレームごとに音声波形の自己相関係数

$$\rho = (\rho_0, \rho_1, \cdots\cdots\cdots\cdots, \rho_p) \qquad (3.1)$$

を求める。VCV音節はρの時系列

$$P = (\rho_1, \rho_2, \cdots\cdots\cdots\cdots, \rho_N) \qquad (3.2)$$

として表現される。ただしNはVCV音節のフレーム数であり，ρ_iは第iフレームの自己相関係数である。この様子を表3.2に示す。

表3.2　自己相関係数による音声の表現

→ フレーム番号

↓相関係数の次数 ＼	1	2	3	·····	j	·····	N
0	ρ_{10}	ρ_{20}	ρ_{30}	·····	ρ_{j0}	·····	ρ_{N0}
1	ρ_{11}	ρ_{21}					
2	ρ_{12}						
⋮	⋮				⋮		⋮
i	ρ_{1i}	·····			ρ_{ij}		
⋮	⋮				⋮		⋮
p	ρ_{1p}	·····			ρ_{jp}	·····	ρ_{Np}

3.2.3 認識系

ＶＣＶ音節を認識するシステムは図3.1に示したような構成になっている。まずスイッチＳ

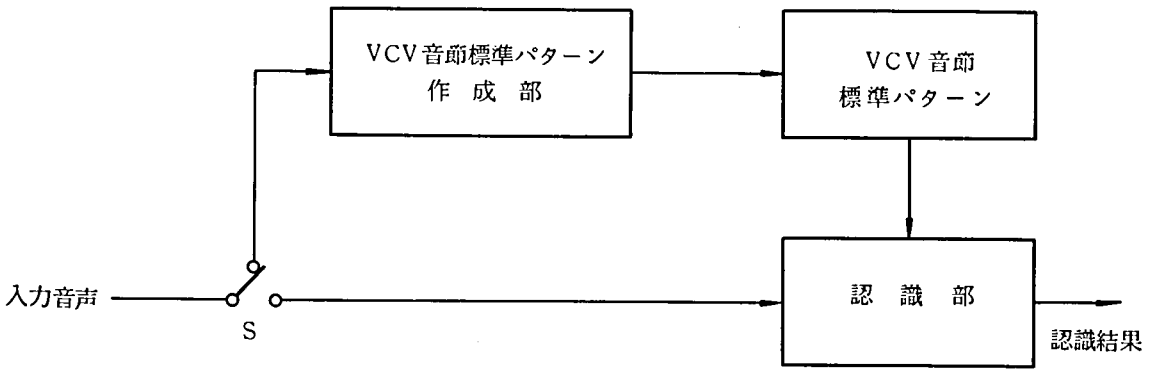


図3.1　ＶＣＶ音節を認識するシステム

を上側に倒し，発声者の発声した音声からＶＣＶ音節の標準パターンを作成する。各ＶＣＶ音節の標準パターンは式（3.2）と同様自己相関係数の時系列

$$\boldsymbol{R}=(\ r_1,\ r_2\ \cdots\cdots\cdots,\ r_M\ ) \qquad (3.3)$$

としてあらわされる。ただしMはフレーム数でありr_jは第jフレームの自己相関係数である。このような$\boldsymbol{R}$はＶＣＶ音節の種類だけ作られ，記憶される。次に認識をおこなう時には，スイッチＳを下側に倒し，発声者は任意のＶＣＶ音節を発声する。入力音声を$\boldsymbol{P}$とすると，認識部では$\boldsymbol{P}$と各ＶＣＶ音節の標準パターン$\boldsymbol{R}$とのパターンマッチングがおこなわれ，それらの類似度が計算される。その結果，類似度が最大であるＶＣＶ音節のカテゴリーに入力が属するものと決定する。標準パターンの作成方法，認識方法については3.3節，3.4節で述べる。

3.3 標準パターン作成法[116]

3.3.1 学習サンプルの非線形な平均

一般に音声パターンは発声した場合ごとにパターンの変動が生じる。したがって未知サンプルに対する適応性のある標準パターンを作るには，同じカテゴリーに属するいくつかの学習サンプルを平均するという方法がとられる。その際考慮しなければならない問題として発声速度

の変動がある。一般に人間の音声は発声した場合ごとに発声速度に変動がある。このような変動は，音声パターンの時間軸方向の伸縮とみなすことができる。しかもその伸縮は非線形なものであることが知られている。(例えば，早く発声するにつれ，まず母音の定常部分が短くなり，ついで，音韻間の遷移が早くおこなわれるようになるという事実がある。) したがって複数個の学習サンプルの平均をおこなう際には，あらかじめ時間軸の非線形な伸縮の正規化をおこなう必要がある。このような正規化を動的計画法（以後略してＤＰと呼ぶ）を用いておこなう方法がいくつか提案されており，[33][34][35] いずれも良い結果が得られたことが報告されている。そこでＤＰを用いて時間軸の正規化を行うことにより，複数個の学習サンプルから標準パターンを作成する。次にその説明をする。

(1) 特定のＶＣＶ音節に注目する。その学習サンプルを

$$\{ \boldsymbol{P}_i = (\boldsymbol{\rho}_1^i, \boldsymbol{\rho}_2^i, \cdots\cdots\cdots\cdots, \boldsymbol{\rho}_{Mi}^i) \} \qquad (3.4)$$

$$(i = 1, 2, \cdots\cdots, R ; R\text{は学習サンプルの個数})$$

とする。これらの学習サンプルはいずれも時間軸の非線形な伸縮を生じているので，まずその補正をおこなう必要がある。その際，正規化の基準となるサンプルが必要である。

これを基準サンプルと呼び

$$\boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots\cdots\cdots\cdots, \boldsymbol{q}_M)$$

であらわす。

(2) 次に各学習サンプルの時間軸の正規化をおこなう。いま，学習サンプル $\boldsymbol{P}_i$ の時間軸の正規化をするには基準サンプル $\boldsymbol{Q}$ に時間軸の変換をおこなう関数 f_i を施して $\boldsymbol{P}_i$ と同じ時間構造を持った次に示すような $\boldsymbol{Q}^i$ を求めればよい。

$$\begin{aligned} \boldsymbol{Q}^i &= (\boldsymbol{q}_1^i, \boldsymbol{q}_2^i, \cdots\cdots\cdots\cdots, \boldsymbol{q}_{Mi}^i) \\ &= (\boldsymbol{q}_{fi(1)}, \boldsymbol{q}_{fi(2)}, \cdots\cdots\cdots\cdots, \boldsymbol{q}_{fi(Mi)}) \end{aligned} \qquad (3.6)$$

ただし fi は次の条件を満足する関数である。

$$\left.\begin{array}{ll} \text{(i)} & f_i(1) = 1, \quad f_i(Mi) = M \\ \text{(ii)} & f_i(j) \geqq f_i(j-1) \quad (j = 2, 3, \cdots\cdots, M_i) \end{array}\right\} \qquad (3.7)$$

(i)はP_iとQの始端と終端が対応することを示している。(ii)はf_iが単調増加関数であることを示しているが，これは音声の時間的な順序関係を保つために明らかに必要な条件である。

さて，時間軸の正規化がどの程度されたかを評価する関数としてはいろいろ考えられるが，P_iとQの対応する要素同士を比較することによる単純なマッチングの度合が最大になったときに時間軸の正規化がされたものと考えるのが自然である。このような考え方から，ρ_j^iとq_kの類似度を

$$l\ (\rho_j^i,\ q_k) \tag{3.8}$$

$$(1 \leqq i \leqq R,\ 1 \leqq j \leqq M_i,\ 1 \leqq k \leqq M)$$

とすると上の問題は

$$L\ (P_i,\ Q,\ f_i) = \sum_{j=1}^{Mi} l\ (\rho_j^i,\ q_{f_i(j)}^{\ i}) = \sum_{j=1}^{Mi} l\ (\rho_j^i,\ q_{f_i(j)}) \tag{3.9}$$

を最大にする関数f_iを求める問題に帰着される。(図3.2)

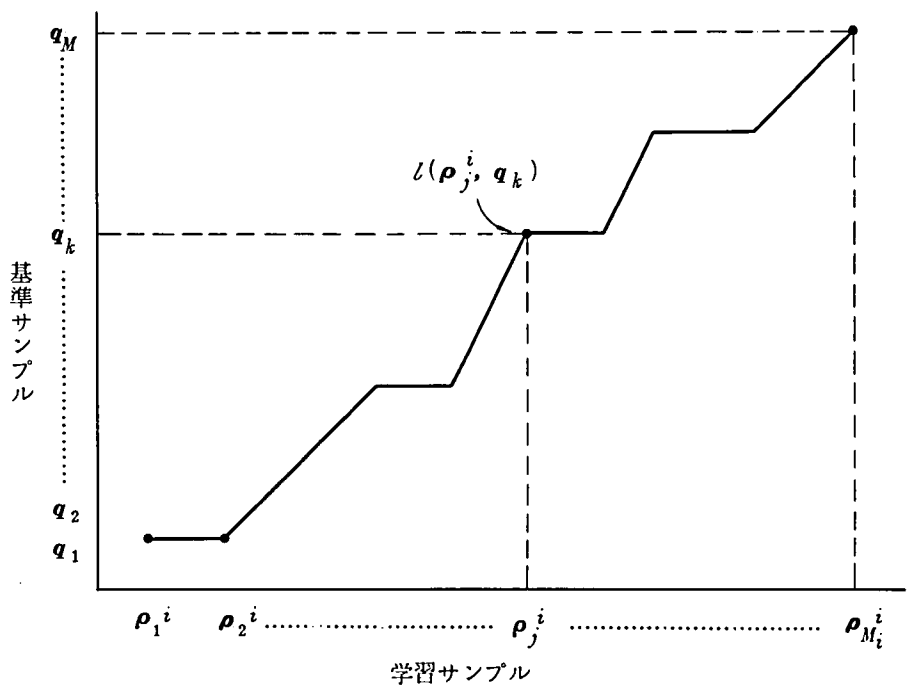


図3.2　時間軸の正規化

この問題はＤＰを用いて簡単にとくことができる。まず，式（3.7）の(ii)を，制限を強くした具体的な条件であらわすことにする。制限を強めることは，本来 $\boldsymbol{P}_i$ と $\boldsymbol{Q}$ の時間構造がそれ程違っているはずがないことからも自然である。ここで用いるＤＰは次の２つの条件のいずれかを用いることにする。

$$\text{(i)}\quad f_i(j) = \begin{cases} f_i(j-1) \\ f_i(j-1)+1 \end{cases} \qquad (3.10)$$

$$\text{(ii)}\quad f_i(j) = \begin{cases} f_i(j-1) \\ f_i(j-1)+1 \\ f_i(j-1)+2 \end{cases} \qquad (3.11)$$

これは具体的にはＤＰをおこなう際，順次加えられる要素間の位置関係が図3.3に示したよ

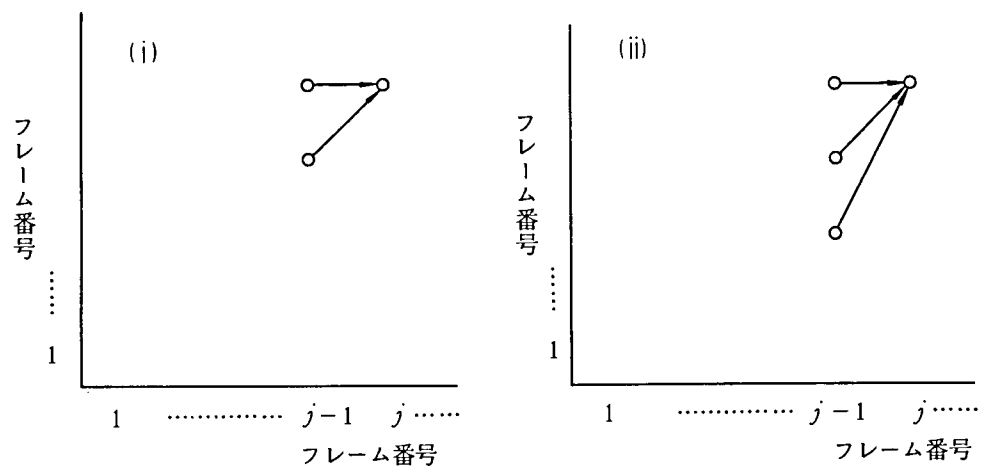


図3.3　ＤＰの計算の際加えられる要素間の関係

うになっていることを示している。いま仮に式（3.11）の条件を採用したとしよう。このとき $\max_{f_i} L(\boldsymbol{P}_i, \boldsymbol{Q}, f_i)$ は次のようにDPによって能率良く求めることができる。

まず，$L_i(j, k)$ を次のように定義する。

$$L_i(j, k) = \max_{f_i} L((\boldsymbol{\rho}_1^i, \boldsymbol{\rho}_2^i, \cdots\cdots, \boldsymbol{\rho}_j^i), (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots, \boldsymbol{q}_k), f_i) \qquad (3.12)$$

これは学習サンプルが $(\boldsymbol{\rho}_1^i, \boldsymbol{\rho}_2^i, \cdots\cdots, \boldsymbol{\rho}_j^i)$，基準サンプルが $(\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots, \boldsymbol{q}_k)$ のときの最大マッチングの値である。$L_i(j, k)$ は

$$L_i(1,\ 1) = l\ (\boldsymbol{\rho}_1^i\ ,\ \boldsymbol{q}_1\ ) \tag{3.13}$$

なる初期条件のもとで次の漸化式

$$L_i(j,\ k) = \max \begin{cases} L_i(j-1,\ k\ \ \ ) + l(\boldsymbol{\rho}_j^i,\ \boldsymbol{q}_k) \\ L_i(j-1,\ k-1) + l(\boldsymbol{\rho}_j^i,\ \boldsymbol{q}_k) \\ L_i(j-1,\ k-2) + l(\boldsymbol{\rho}_j^i,\ \boldsymbol{q}_k) \end{cases} \tag{3.14}$$

を順次とくことによって求められる。

このとき，

$$L_i(M_i,\ M) = \max_{f_i} L\ (\boldsymbol{P}_i\ ,\ \boldsymbol{Q},\ f_i\ ) \tag{3.15}$$

となる。また，$L_i(M_i\ ,\ M)$ を求める際加えた $l\ (\boldsymbol{\rho}_j^i,\ \boldsymbol{q}_k)$ を逆にたどることにより f_iを求めることができる。

(3) (2)の操作をすべての $\boldsymbol{P}_i$ についておこなうと関数の集合 $\{f_i\}$ ($i = 1, 2, \cdots\cdots, R$) が得られる。各 f_iは，$\boldsymbol{P}_i$ 中の各特徴ベクトルを $\boldsymbol{Q}$ の特徴ベクトルに対応づける写像関数と考えることができる。(図 3.4) いま，$\boldsymbol{q}_k$ に対応づけられた特徴ベクトルの集合を $P(k)$ とする。すなわち $P(k)$は次式であらわすことができる。

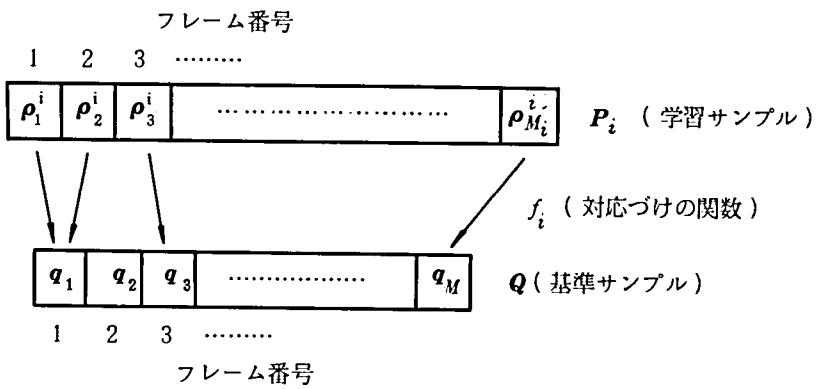


図 3.4　f_i による $\boldsymbol{P}_i$ と $\boldsymbol{Q}$ の対応づけ

$$P(k) = (\boldsymbol{\rho}_j^i \mid f_i(j) = k,\ \ 1 \leqq j \leqq M_i\ ,\ \ i = 1, 2, \cdots\cdots, R) \tag{3.16}$$

$$(k = 1, 2, \cdots\cdots, M)$$

$P(k)$に含まれる特徴ベクトルを平均して $\boldsymbol{r}_k$ を得る。すなわち

$$r_k = \frac{\sum_{\rho_j^i \in P(k)} \rho_j^i}{|P(k)|} \qquad (3.17)$$

（$|P(k)|$ は集合 $P(k)$ の要素の数）

この結果得られる M 個のベクトルの系列を $\boldsymbol{R}$ とする。

$$\boldsymbol{R} = (r_1, r_2, \cdots\cdots, r_M) \qquad (3.18)$$

このようにして作られた $\boldsymbol{R}$ を標準パターンとする。これはいくつかの学習サンプルの時間構造のちがいを補正したのち平均して得られたものであるから，標準パターンとしての条件を満足しているといえる。

3.3.2 種々の標準パターン

本実験では，3.1 で述べた標準パターン作成法を基本として大きくわけて 5 種類の標準パターンを作成した。これは，他の音声単位（例えば音韻等）にもとづいた標準パターンとの比較実験をおこなったり，種々条件をかえて認識率の高い標準パターンを得るための実験をおこなったりするためである。次に，各標準パターンについて説明する。

3.3.2.1 標準パターンA

これは，音韻を音声単位とした標準パターンであり，ＶＣＶ音節を音声単位とした標準パターンとの比較実験に用いる。具体的な作成法は次の通りである。

(1) 母音，子音を含んだ連続音声中から視察により母音区間，子音区間を切り出す。

(2) (1)の操作により切り出された音韻 / x / の各フレームの相関係数を ρ_1^x, ρ_2^x, ……, ρ_M^x とすると，それらを平均してできる。

$$\rho^x = \frac{1}{M} \sum_{i=1}^{M} \rho_i^x \qquad (3.19)$$

を音韻 / x / の標準パターンとする。

(3) ＶＣＶ音節の標準パターンとの比較をおこないやすくするため，音韻の標準パターンを組

み合わせて見かけ上のＶＣＶ音節の標準パターンを作る。すなわち，$/V_1C_2V_3/$ なるＶＣＶ音節の標準パターンを $P_{V_1C_2V_3}$ であらわすと

$$P_{V_1C_2V_3} = (\rho^{V_1}, \rho^{C_2}, \rho^{V_3}) \qquad (3.20)$$

である。

3.3.2.2 標準パターンB

これは，ＶＣ，ＣＶ音節を音声単位とした標準パターンである。標準パターンＡと同じく，ＶＣＶ音節を音声単位とした標準パターンとの比較実験に用いる。作成法を次に示す。

(1) 学習サンプル……一組のＶＣＶ音節のサンプルを用意する。各サンプルを音声区間の中央で２つの区間に分け，２個のＶＣ，ＣＶ音節のサンプルを作る。（図3.5） すべてのサンプル

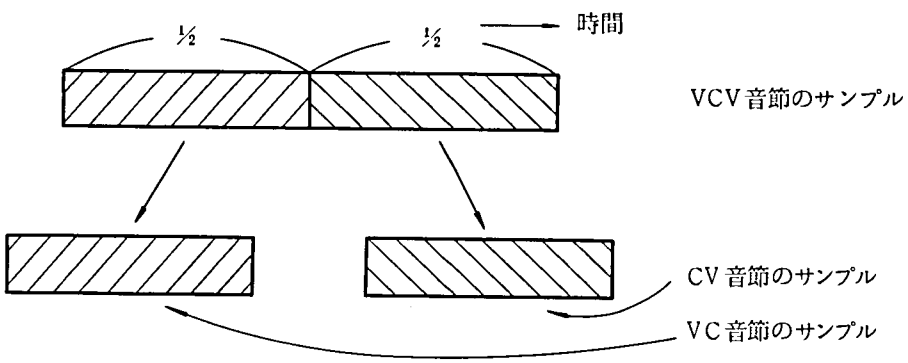


図3.5 ＶＣＶ音節のサンプルからＶＣ，ＣＶ音節のサンプルを作る方法

についてこの操作をおこなうと，各ＶＣ，ＣＶの音節のカテゴリーには５個ずつのサンプルが属することになる。これを学習サンプルとする。

(2) 基準サンプル……(1)と同じ操作で作ったＶＣ，ＣＶ音節のサンプルを各カテゴリーにつき１個ずつ用意する。用意された各サンプルから２フレームおきに抽出したフレームの系列を基準サンプルとする。（図3.6）

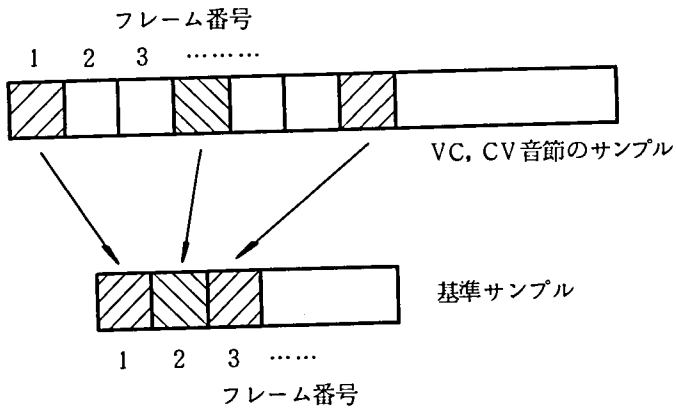


図3.6　ＶＣ，ＣＶ音節の基準サンプルの作成法

(3)　以上の学習サンプルと基準サンプルを用いて3.3.1で述べた方法によりＶＣ，ＣＶ音節の標準パターンを作る。例えば $/V_1C_2V_3/$ なるＶＣＶ音節の標準パターンが必要な時は $/V_1C_2/$, $/C_2V_3/$ なるＶＣ，ＣＶ音節の標準パターンを用意する。これらの標準パターンを

$$\boldsymbol{P}_{V_1C_2} = (\boldsymbol{\rho}_1^1, \boldsymbol{\rho}_2^1, \cdots\cdots, \boldsymbol{\rho}_{M_1}^1) \qquad (3.21)$$

$$\boldsymbol{P}_{C_2V_3} = (\boldsymbol{\rho}_1^2, \boldsymbol{\rho}_2^2, \cdots\cdots, \boldsymbol{\rho}_{M_2}^2) \qquad (3.22)$$

とすると $/V_1C_2V_3/$ 音節の標準パターンは次式で与えられる。

$$\boldsymbol{P}_{V_1C_2V_3} = (\boldsymbol{\rho}_1^1, \boldsymbol{\rho}_2^1, \cdots\cdots, \boldsymbol{\rho}_{M_1}^1, \boldsymbol{\rho}_1^2, \boldsymbol{\rho}_2^2, \cdots\cdots, \boldsymbol{\rho}_{M_2}^2) \qquad (3.23)$$

3.3.2.3　標準パターンC

ＶＣＶ音節のサンプルをそのまま標準パターンとして用いるものである。これは，パターンの変動の平均化がされていない場合，認識結果にどのような影響を及ぼすかを知るために用いる。具体的には次の2種類を用いる。

(1)　標準パターンC_1……もとのＶＣＶ音節のサンプルから2フレームおきに取りだしたフレームの系列を標準パターンとする。すなわちもとのサンプルを

$$\boldsymbol{P} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \cdots\cdots, \boldsymbol{\rho}_N) \qquad (3.24)$$

とすると標準パターン $\boldsymbol{P}_S$ は次式で与えられる。(図3.7(a))

$$P_S = (\rho_1, \rho_4, \cdots\cdots\cdots, \rho_{3M-1}) \qquad (3.25)$$

（ただし　$M = [\frac{N+1}{3}]$，〔 〕はガウス記号）

(2)　標準パターンC₂……もとのＶＣＶ音節のサンプルの中央部分１/２を取り出したものを標準パターンとする。すなわちもとのサンプルが式（3.24）で与えられたとき標準パターンP_sは次式で与えられる。（図3.7 (b)）

$$P_S = (\rho_{[\frac{N}{4}]}, \rho_{[\frac{N}{4}]+1}, \cdots\cdots\cdots, \rho_{[\frac{3N}{4}]}) \qquad (3.26)$$

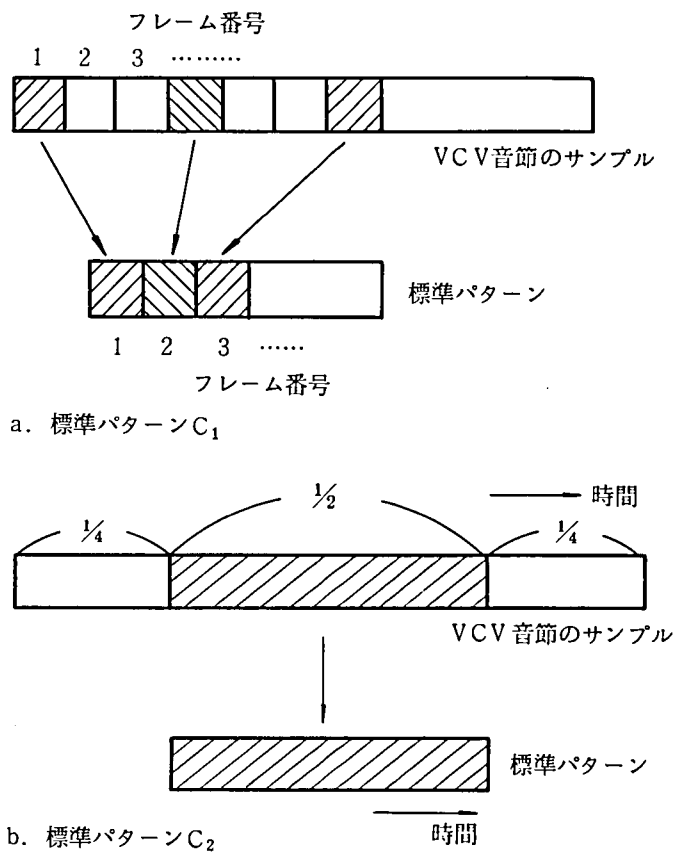


図3.7　標準パターンＣの作成法

3.3.2.4 標準パターンD

5組の学習サンプルから，3.3.1で述べた方法に従って作った標準パターンである。具体的には次の3種類を用いる。

(1) 標準パターンD_1

○学習サンプル……各ＶＣＶ音節ごとに5個のサンプルを用意し学習サンプルとする。

○基準サンプル……学習サンプルのうち1組をとりだし，各サンプルのフレームの系列から2フレームおきにとりだしたフレームの系列を基準サンプルとする。

以上の学習サンプルと基準サンプルから3.3.1の方法で標準パターンを作成する。

(2) 標準パターンD_2

標準パターンD_1より

$$\left[\frac{2k-1}{10}M\right] \qquad (3.27)$$

（Mは標準パターンのフレーム数；$k=1, 2, \cdots\cdots, 5$）

番目のフレームを順次とりだしてできる5個のフレーム系列を改めて標準パターンD_2とする。（図3.8）

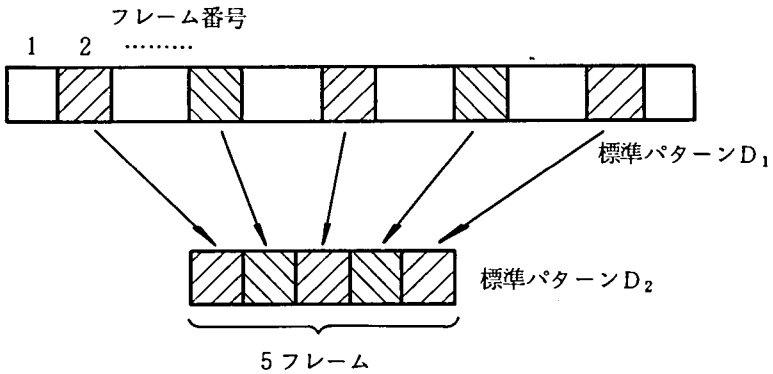


図3.8　標準パターンD_2の作成法

(3) 標準パターンD_3

標準パターンD_2と同様，標準パターンD_1より3個のフレームをぬきだしてできるフレームの系列を標準パターンD_3とする。

このような3種類の標準パターンを用いたのは，標準パターンがスペクトルパターンの時間

推移に関する情報をより多く持つことが認識率にどの程度寄与するかを調べるためである。

3.3.2.5 標準パターンE

標準パターンDが5組の学習サンプルから作成されるのに対し，標準パターンEは10組の学習サンプルから作った標準パターンである。基準サンプルの作り方によって3種類に分けられる。

(1) 標準パターンE_1……学習サンプルのうち一組をそのまま基準サンプルとする。

(2) 標準パターンE_2……3.3.1で述べた標準パターンの作成法から明らかなように，各学習サンプルの時間正規化は基準サンプルを基準としておこなわれる。したがって，特定の音声サンプルをそのまま基準サンプルとすると，標準パターンの時間構造が基準サンプルとして用いた音声サンプルの時間構造に似てくることになる。本来すべての未知パターンに対する適応性をもつべき標準パターンがこのように特定のサンプルに似た時間構造を持つことは望ましくない。この問題を解決する1つの方法として各学習サンプルをそのまま（すなわち時間軸の正規化をおこなわずに）平均して基準サンプルとすることにする。

いま，特定のVCV音節に注目し，その学習サンプルを

$$\{ \boldsymbol{P}_i = (\boldsymbol{\rho}_1^i, \boldsymbol{\rho}_2^i, \cdots\cdots, \boldsymbol{\rho}_{M_i}^i) \} \qquad (3.28)$$

$$(i = 1, 2, \cdots\cdots, 10)$$

とすると基準サンプル$\boldsymbol{Q}$は次式で与えられる。（図3.9）

$$\boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots, \boldsymbol{q}_M) \qquad (3.29)$$

ただし $\quad \boldsymbol{q}_k = \frac{1}{10} \sum_{i=1}^{10} \boldsymbol{\rho}_k^i$

$$(k = 1, 2, \cdots\cdots, M; \quad M = \min(M_1, M_2, \cdots\cdots, M_{10}))$$

この$\boldsymbol{Q}$を基準パターンとして作った標準パターンが標準パターンE_2である。

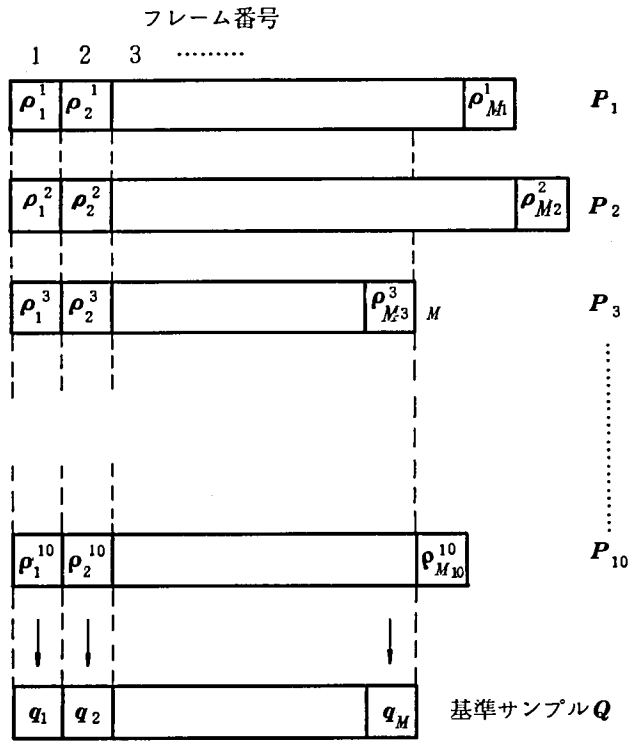


図 3.9　標準パターン E_2 の基準サンプルの作成法

(3)　標準パターン E_3 ……式（3.29）の Q は学習サンプルを単純平均したものであり，スペクトルパターンの時間推移に関する情報がかなり失われていると思われる。したがって，Q を基準サンプルとして作った標準パターン E_2 はまだ十分満足すべきものではない。この対策として E_2 を基準サンプルとしてもう一度標準パターンを作りなおすことが考えられる。この方法で作った標準パターンを E_3 とする。

以上説明した標準パターンの相違をわかりやすくするために特徴をまとめて表 3.3 に示す。

表3.3　標準パターンの分類

標準パターンの種類		音声単位	基準サンプル	学習サンプル数	学習サンプルの平均化法	DP	フレーム数
A		音韻			単純平均		3フレーム
B		VC，CV音節	1組	5組	非線形平均	式（3.10）	全フレームの⅓
C	C_1	VCV音節	〃	1組	平均しない		〃
	C_2	〃	〃	〃	〃		全フレームの½
D	D_1	〃	〃	5組	非線形平均	式（3.10）	全フレームの⅓
	D_2	〃	〃	〃	〃	〃	5フレーム
	D_3	〃	〃	〃	〃	〃	3フレーム
E	E_1	〃	〃	10組	〃	式（3.11）	全フレームの⅓
	E_2	〃	10組の単純平均	〃	〃	〃	〃
	E_3	〃	標準パターンE_2	〃	〃	〃	〃

3.4　認識法[116]

3.4.1　時間軸の正規化を行ったパターンマッチング法

3.3.1で述べたように，音声は発声した場合ごとに時間軸の非線形な伸縮が生じる。したがって，入力音声と標準パターンとのパターンマッチングをおこなう際にも時間軸の正規化をおこなう必要がある。時間軸の正規化法については3.3.1でくわしく述べたのでここでは簡単に説明する。

入力音声を

$$\boldsymbol{P}=(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\cdots\cdots\cdots,\boldsymbol{\rho}_N) \qquad (3.30)$$

標準パターンを

$$\boldsymbol{R}=(\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots\cdots\cdots,\boldsymbol{r}_M) \qquad (3.31)$$

とする。$\boldsymbol{R}$の時間軸の変換をおこなう関数をfとし，変換の結果得られた標準パターンを$\boldsymbol{R}'$とすると

$$\boldsymbol{R}' = (\boldsymbol{r}'_1, \boldsymbol{r}'_2, \cdots\cdots, \boldsymbol{r}'_N) = (\boldsymbol{r}_{f(1)}, \boldsymbol{r}_{f(2)}, \cdots\cdots, \boldsymbol{r}_{f(N)}) \tag{3.32}$$

である。ただしfは次の条件を満足する。

$$\left.\begin{array}{ll} \text{(i)} & f(1)=1, \quad f(N)=M \\ \text{(ii)} & f(i) \geqq f(i-1), \quad (i=2, 3, \cdots\cdots, N) \end{array}\right\} \tag{3.33}$$

$\boldsymbol{\rho}_i$と$\boldsymbol{r}_j$の類似度を

$$l(\boldsymbol{\rho}_i, \boldsymbol{r}_j) \tag{3.34}$$

$$(1 \leqq i \leqq N, \quad 1 \leqq j \leqq M)$$

とするとfは

$$L(\boldsymbol{P}, \boldsymbol{R}, f) = \sum_{i=1}^{N} l(\boldsymbol{\rho}_i, \boldsymbol{r}'_i)$$

$$= \sum_{i=1}^{N} l(\boldsymbol{\rho}_i, \boldsymbol{r}_{f(i)}) \tag{3.35}$$

の最大値を与える解として求めることができる。また式（3.35）の値が$\boldsymbol{P}$と$\boldsymbol{R}$の最大マッチングの値となる。式（3.35）はＤＰを用いて簡単に求めることができる。実際にＤＰの計算をおこなう時には，fにはより強い制限として次のいずれかの条件をつける。

$$\text{(i)} \quad f(i) = \begin{cases} f(i-1) \\ \\ f(i-1)+1 \end{cases} \tag{3.36}$$

$$\text{(ii)} \quad f(i) = \begin{cases} f(i-1) \\ \\ f(i-1)+1 \\ \\ f(i-1)+2 \end{cases} \tag{3.37}$$

これらの条件は標準パターン作成法で述べた式（3.10），式（3.11）に対応するものである。

3.4.2 種々の制限

実際の認識実験をおこなう際には，認識率を上げるため3.4.1で述べた認識法に種々の制限を加える。

3.4.2.1 パワー情報の使用

自己相関係数で表現された音声は，スペクトルパターンに関する情報しか持っておらず，その他の重要な情報，たとえばパワー情報とか，ピッチ情報が含まれていない。したがって有声音と無声音の判別の際には誤りがおこりやすいと考えられる。そこで有声・無声の判別をおこなうため，最も簡単な情報としてパワー情報を採用する。一般に，無声子音の前には無音区間が生じる。この様子を調べるため，2名の発声者がそれぞれ1組ずつ発声したＶＣＶ音節について無音区間のフレーム数の分布を求めて表3.4に示す。この表をもとにして図3.10に示す判定論理を作った。したがってパワー情報を認識に用いる際には，まず図3.10の判定論理で子音の分類をした後，3.4.1の方法で認識するという方法をとる。

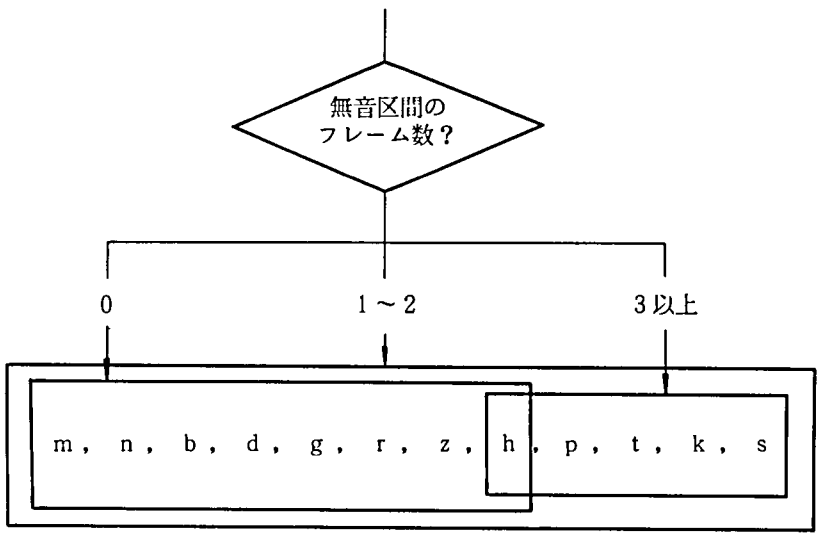


図3.10　子音を分類する判定論理

表3.4　無音区間のフレーム数の分布

子音	発声者	無音区間のフレーム数									
		0	1	2	3	4	5	6	7	8	9
m	N	24	1								
	K	25									
n	N	25									
	K	25									
b	N	25									
	K	25									
d	N	25									
	K	24	1								
g	N	25									
	K	25									
r	N	25									
	K	25									
z	N	25									
	K	25									
p	N					10	11	4			
	K			1	4	11	8	1			
t	N				6	6	3	8	1	1	
	K			2	4	10	2	3	3	1	
k	N		1	5	13	2	1	1	2		
	K				4	9	8	4			
s	N					8	8	5	3		
	K				2	6	9	7	1		
h	N	21		2	1	1					
	K	12	3	4	2	4					

3.4.2.2　類似度を計算する際の重みづけ

入力のＶＣＶ音節と標準パターンとの類似度を計算する際，3.4.1の方法では両端の母音部分も平等に類似度を加えている。しかしながら実際にはＶＣＶ音節の各部分が平等に認識に寄与しているとは考えられない。したがって，認識の際の有効性に応じて各部分の類似度に重み

づけをして類似度和を求めるのが有効と考えられる。ここでは，ＶＣＶ音節の認識には，両端の母音部分はあまり寄与しておらず，中央の子音部分の微妙なスペクトルパターンの差が認識に大きく影響しているという仮定のもとに次に示すような台形関数を重み関数として採用する。

$$\begin{cases} f(x)=0 \\ \qquad (1 \leqq x < [\frac{N}{3}],\ [\frac{2N}{3}] < x \leqq N) \\ f(x)=1 \quad ([\frac{N}{3}] \leqq x \leqq [\frac{2N}{3}]) \end{cases} \qquad (3.38)$$

（Nは入力のフレーム数）

3.4.2.3、継続時間の制限

3.4.1では入力音声と標準パターンの時間構造があまり違わないという仮定のもとに，入力音声の各部分と標準パターンの各部分の対応づけをおこなう際，式（3.36）または式（3.37）のような制限をつけた。しかし，これだけの制限であれば標準パターンの１フレームには入力音声の何フレームでも対応づけられることになり実際の物理現象との対応上不自然である。したがって，認識のためのＤＰを計算する際，標準パターンの１フレームに対応づけられる入力音声のフレーム数に上限を与えておくことが望ましい。実際にこの方法を用いる時には上限を３フレームとした。

3.4.2.4 cut off frequency の変更

3.2.2で述べたように，入力音声は低域通過フィルターを通して3.2 kHz 以上の周波数は切ってある。しかしながら子音の中には3.2 kHz 以上の高域にエネルギーが集中しているものがあり，3.2 kHz 以上の周波数を切ることは当然このような子音の認識には不利と考えられる。そこで，認識率を上げるための１つの実験としてcut off frequency を6 kHz にした実験をおこなった。この場合，スペクトルパターンの周波数軸が２倍に伸びることになるから当然スペクトルパターンは複雑になる。そこでスペクトルパターンの微細な構造を表現するため式（3.1）のPを15とした。

3.5 認識実験

3.5.1 実験に用いた音声サンプル

標準パターン作成用の学習サンプルや認識の際の未知サンプルとして1名の発声者による14組の音声サンプルを用意した。これらの音声サンプルをそれぞれサンプルa，サンプルb，…サンプルnと呼ぶことにする。各サンプルについて簡単に説明する。

(1) サンプルa

/ ma mi mu me mo /，/ na ni nu ne no /，…… ，/ za zi zu ze zo /なる連続音声。標準パターンAを作成するのに用いる。

(2) サンプルb，c

子音を有声子音（/ m, n, b, d, g, r, z /）に限定した175個のＶＣＶ音節の音声サンプル2組。

(3) サンプルd～サンプルn

300種類のＶＣＶ音節の音声サンプル11組。

3.5.2 実験の説明

標準パターン，認識法を種々組みあわせて計20種類の実験をおこなった。これを実験1，実験2，……，実験20と呼ぶことにする。各実験の比較がおこないやすいよう表3.5に実験の分類をしておいた。次に各実験について簡単に説明する。

表 3.5 実 験 の 分 類

実験番号	標準パターンの種類	認識対象	サンプル	DP	認識の際の制限	実験に用いたサンプル			認識率
						基準サンプル	学習サンプル	未知サンプル	
1	A	175種類のVCV音節	未知サンプル	(3.36)			a	b	44.0%
2	B	〃	〃	〃		b	c	b	60.0%
3	B	〃	〃	〃		b	b	c	56.6%
4	B	〃	学習サンプル	〃		c	c	c	97.7%
5	B *	〃	〃	〃	*実験2, 3の標準パターンを平均したもの	b	b, c	b	94.9%
6	B *	〃	〃	〃		b	b, c	c	90.3%
7	C_1	〃	未知サンプル	〃			c	b	49.7%
8	C_2	〃	〃	〃			c	b	40.6%
9	D_1	〃	〃	〃		d	d～h	n	69.1%
10	D_1	300種類のVCV音節	〃	〃		〃	〃	i	71.7%
11	D_2	〃	〃	〃		〃	〃	〃	54.3%
12	D_3	〃	〃	〃		〃	〃	〃	43.7%
13	E_1	〃	〃	(3.37)	母音既知	〃	d～m	n	78.3%
14	E_2	〃	〃	〃	〃	d～m	〃	〃	81.3%
15	E_2	〃	〃	〃	〃, power 情報使用	〃	〃	〃	88.0%
16	E_2	〃	学習サンプル	〃	〃, 〃	〃	〃	d	91.0%
17	E_2	〃	未知サンプル	〃	〃, 重みづけ	〃	〃	n	70.3%
18	E_2	〃	〃	〃	〃, 継続時間制限	〃	〃	〃	80.3%
19	E_2	175種類のVCV音節	〃	〃	〃, cut off frequency 6 KHz	〃	〃	〃	89.1%
20	E_3	300種類のVCV音節	〃	〃	〃, power 情報使用	〃	〃	〃	88.3%

(1) 標準パターンAを用いた実験（実験1）

標準パターンAは音韻を音声単位とした標準パターンである。この実験は，ＶＣＶ音節を音声単位にとった標準パターンとの比較をするために行った。

(2) 標準パターンBを用いた実験（実験2～6）

標準パターンBはＶＣ，ＣＶ音節を音声単位とした標準パターンである。(1)と同様，ＶＣＶ音節パターンとの比較をするために行った実験である。実験2，3は基準サンプルの選び方が認識にどのような影響をするかを調べるための実験である。実験4では学習サンプルの認識をおこなった。実験5，6で用いた標準パターンは，実験2，3の標準パターンを平均したものである。

すなわち，実験5，6は標準パターンを作成するための学習サンプルの組数をふやすと認識にどのように影響するかを調べるための実験である。

(3) 標準パターンCを用いた実験（実験7，8）

標準パターンCはＶＣＶ音節を音声単位とした標準パターンである。ただしパターンの変動に対する平均化はされていない。したがって，この実験はパターンの変動に対する平均化が認識にどのように影響するかを調べるためのものである。C_1，C_2の2つの標準パターンを用いたのは，ＶＣＶ音節のどの部分が認識に有効なのかを調べるためである。なお，標準パターンA，B，Cを用いた実験ではいずれも認識対象は有声子音のみよりなる175種類のＶＣＶ音節である。

(4) 標準パターンDを用いた実験（実験9～12）

標準パターンDはＶＣＶ音節を音声単位とし，5組の学習サンプルより作成した標準パターンである。実験9は標準パターンA，B，Cとの比較をおこなうための実験であり，認識対象は175種類のＶＣＶ音節である。これ以後の実験では子音は無声子音も含めることとし，300種類のＶＣＶ音節を認識対象とする。（ただし実験19は例外）実験10, 11, 12は標準パターンのフレーム数と認識率の関係を調べるためにおこなった。

(5) 標準パターンEを用いた実験（実験13～20）

標準パターンEはＶＣＶ音節を音声単位とし，10組の学習サンプルより作成した標準パター

ンである。なお，標準パターンＥはフレーム数が多くなり（20フレーム～30フレーム）認識に時間がかかるため，入力音声の母音部分はあらかじめわかっているものとして認識実験をおこなった。後で述べるように，ＶＣＶ音節の母音部分はほぼ完全に認識できるため，このような方法で得られた認識率はそのままＶＣＶ音節の正しい認識率と考えてよいであろう。実験13，14は基準サンプルの作成法がことなる２種類の標準パターンE_1,E_2を用いた比較実験である。実験15以後は認識の際，種々の工夫をして認識率の向上をはかった実験である。まず実験15，16はパワー情報を使った認識実験である。実験15では未知サンプルを，実験16では学習サンプルを認識した。実験17は3.4.2.2で述べた重みづけをおこなったものである。実験18は3.4.2.3で述べたように標準パターンの１フレームに対応づけられる入力音声のフレーム数に制限をつけた認識実験である。実験19は3.4.2.4で述べたようにcut off frequencyを6 kHzにした実験である。ただしこの実験では認識対象は有声子音よりなる175種類のＶＣＶ音節に限定した。最後に実験20は標準パターンE_3を用いた認識実験である。各実験の結果は付録１，表Ａ1.1～Ａ1.22に各実験ごとにまとめてある。

付録の各表は次のような構成になっている。

ａ．ＶＣＶ音節の母音部分の認識率およびconfusion matrix

ｂ．ＶＣＶ音節の子音部分の認識率およびconfusion matrix

ｃ．正しい答の類以度が上位から何番目に分布しているかを示す表。この表ではＶＣＶ音節別に分類してある。

なお，子音列の認識率では（m, n），（b, d, g），（p, t, k）のように調音形式の似た子音をまとめて考えた場合の認識率もあわせてあげておいた。また，3.6の検討ではＣの表に示してあるパーセンテージの累積値と順位の関係をグラフにしたものをしばしば用いるので，以後このようなグラフを「順位と認識率のグラフ」と呼ぶことにする。このグラフで立ち上りの早い曲線であらわされる実験では正しい答の類似度が上位に集中していることを示している。したがって，単に認識率の比較をおこなうより，順位と認識率のグラフを比較した方が正しい判断を下せると思われる。

3.6 検　討

3.5で述べた実験結果にもとづいて種々の検討をおこなう。

3.6.1 標準パターン作成法の検討

(1) 音声単位

音声単位としては，音韻，ＶＣ，ＣＶ音節，ＶＣＶ音節の３種類をとり実験をおこなった。これらを最も公平に比較するため，実験1, 3, 9の結果を比較する。これらの実験で用いた標準パターンはいずれもパターンの変動に対する平均化がされおり，かつ，認識対象も等しいなどできるだけ他の条件は等しくしてある。図3.11に順位と認識率のグラフを，図3.12に子音別の認識率を示す。図3.11から，実験１→実験３→実験９の順に認識率が良くなっていくことがわかる。特に，実験１のグラフは他の２つのグラフに比較して立ち上がりが悪いことは，正しい答

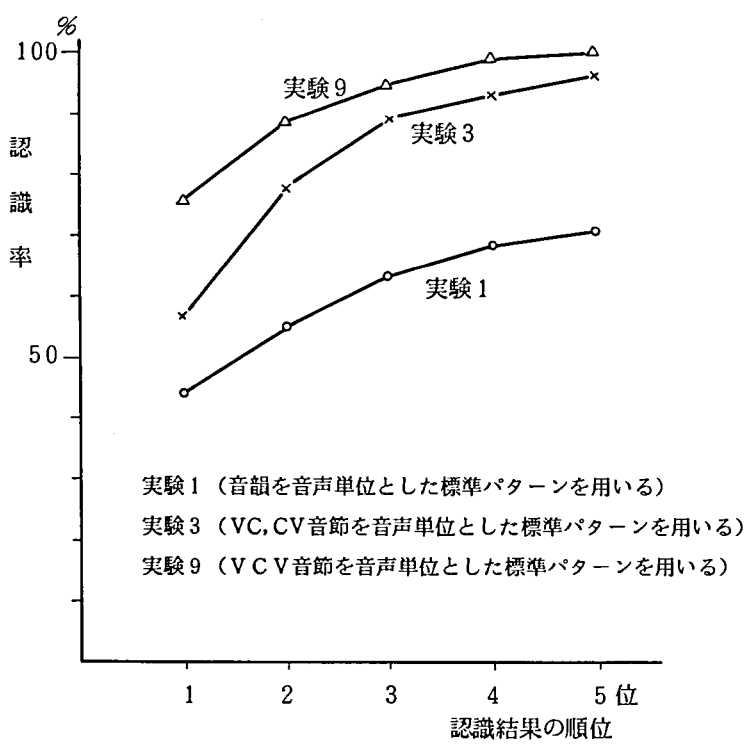


図3.11　順位と認識率のグラフ（音声単位の選び方と認識率の関係）

の類似度が必ずしも上位に存在せず，したがって認識が不安定なことを示している。このことは図 3.12からもわかる。すなわち，実験 1 の結果は子音別の認識率に大きなバラつきがあり認識が不安定なことを示している。

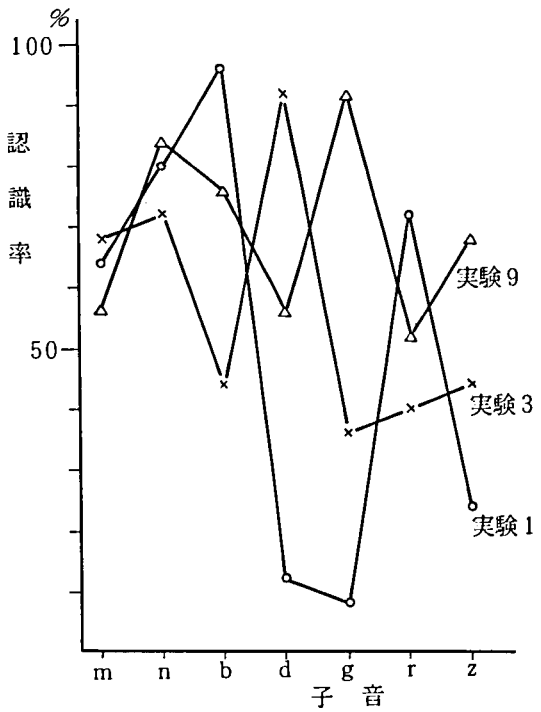


図 3.12　子音別の認識率（音声単位の選び方と認識率の関係）

以上の結果より，認識の面から考えた場合，音声単位としては音韻，ＶＣ，ＣＶ音節，ＶＣＶ音節の中でＶＣＶ音節が最も良いことがわかる。ＶＣ，ＣＶ音節がこれにつぐ。音韻を音声単位とした認識はこれらにくらべると不安定である。

(2)　標準パターンのフレーム数

標準パターン D_2，D_3 は標準パターン D_1 から，それぞれ 5 フレーム，3 フレームを抽出して作った標準パターンである。したがって，標準パターン $D_1 \rightarrow D_2 \rightarrow D_3$ の順にスペクトルパターンの時間推移に関する情報が失われていることになる。スペクトルパターンの時間推移に関する情報が認識に及ぼす影響を調べるため，標準パターン D_1，D_2，D_3 を用いた実験 10, 11, 12 の比較をおこなう。図 3.13に順位と認識率のグラフを，図 3.14に子音別の認識率をグラフにして示す。図 3.13から，フレーム数の増加と共に認識結果が向上するのが見られる。また図 3.14

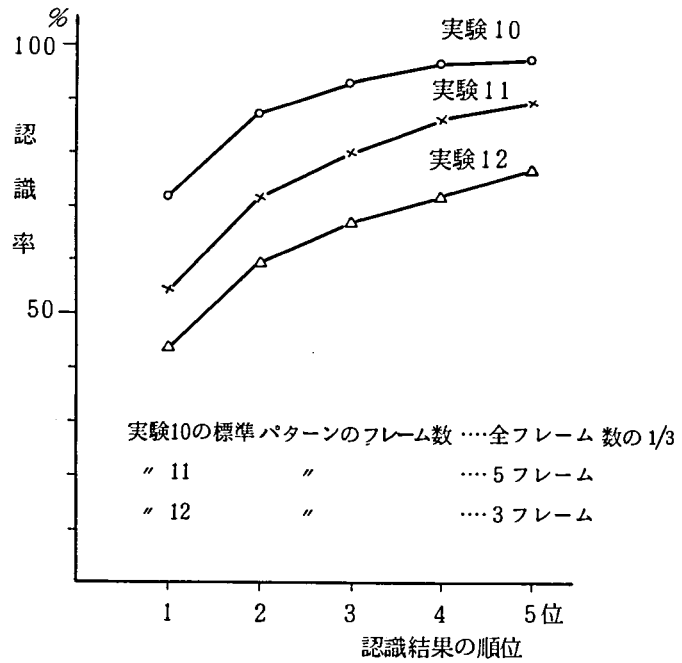


図 3.13　順位と認識率のグラフ
（標準パターンのフレーム数と認識率の関係）

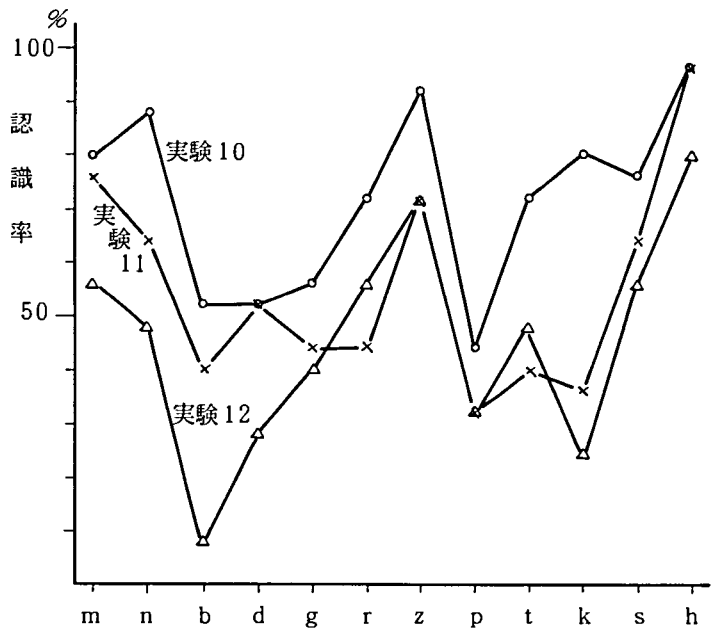


図 3.14　子音別の認識率
（標準パターンのフレーム数と認識率の関係）

においても，数カ所の例外を除くと，子音別の認識率もフレーム数の増加と共に上昇しているのがわかる。これらのことから，標準パターンのフレーム数は多い方が望ましいといえる。すなわち，標準パターンはスペクトルパターンの時間推移に関する情報を多く持っていた方が高い認識率が得られる。

(3) 学習サンプルの組数

標準パターンを作るための学習サンプルの組数が認識に及ぼす影響を調べる。まず学習サンプルが1組の場合（この時はパターンの変動に対する平均化はされない）と5組の場合とを比較するため，実験7と9の結果をとりあげる。図3.15に順位と認識率のグラフを，図3.16に子音別の認識率を示す。いずれのグラフからも，実験9の結果が良いことは明らかである。次に，標準パターンが5組の場合と10組の場合を，実験10と13でおこなう。図3.17に順位と認識率のグラフを図3.18に子音別の認識率を示す。図3.17から実験13の結果の方が良いことがわかるが，それは図3.18を見ると，/ b, d, g /の認識率が上がったことに原因がある。これは，実験10で多かった/ b, d, g / → / p, t, k /なるconfusionが実験13では減少していることによる（付録1参照）。

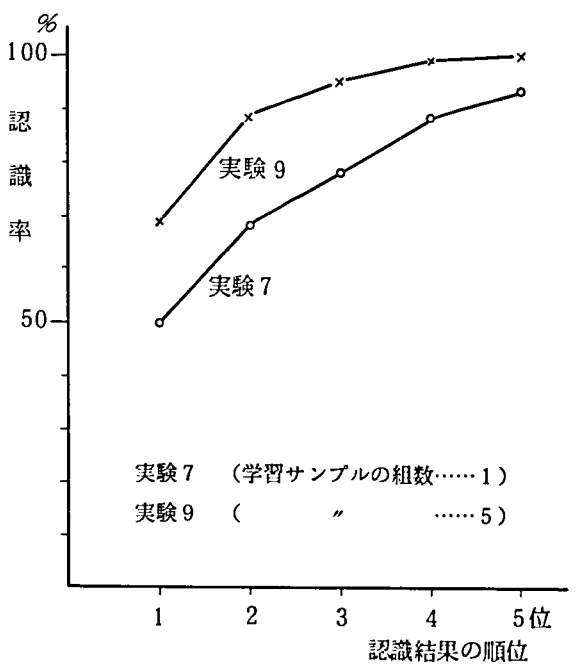


図3.15　順位と認識率のグラフ

（学習サンプルの組数と認識率の関係）

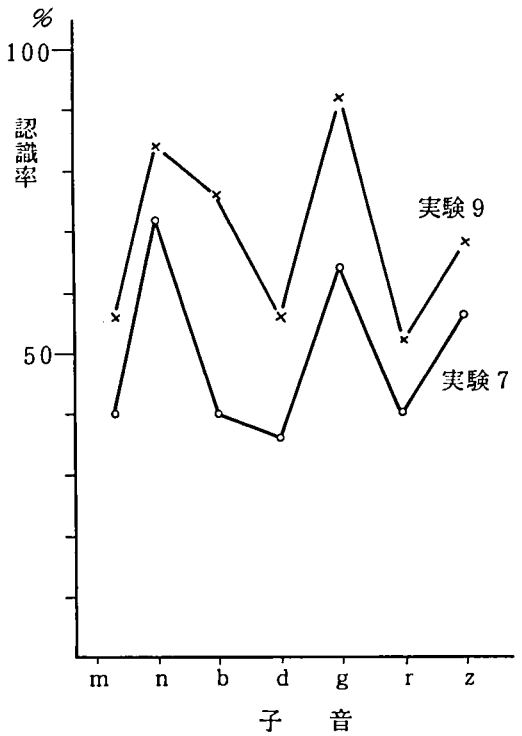


図 3.16　子音別の認識率
（学習サンプルの組数と認識率の関係）

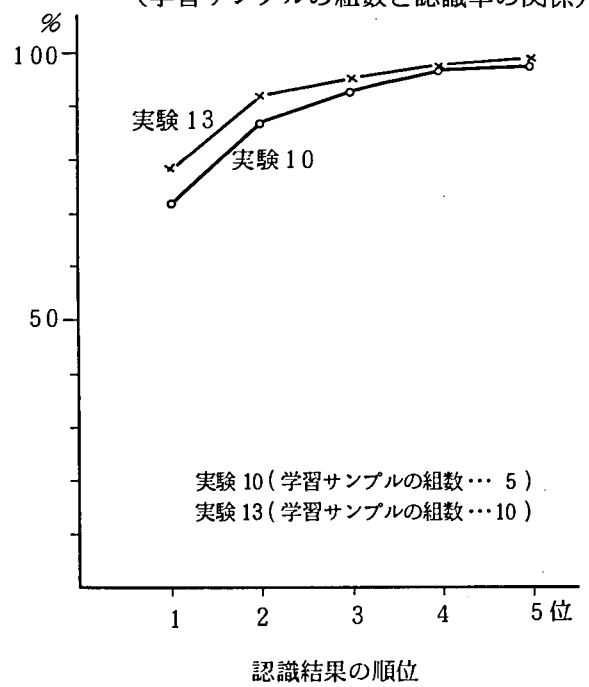


図 3.17　順位と認識率のグラフ
（学習サンプルの組数と認識率の関係）

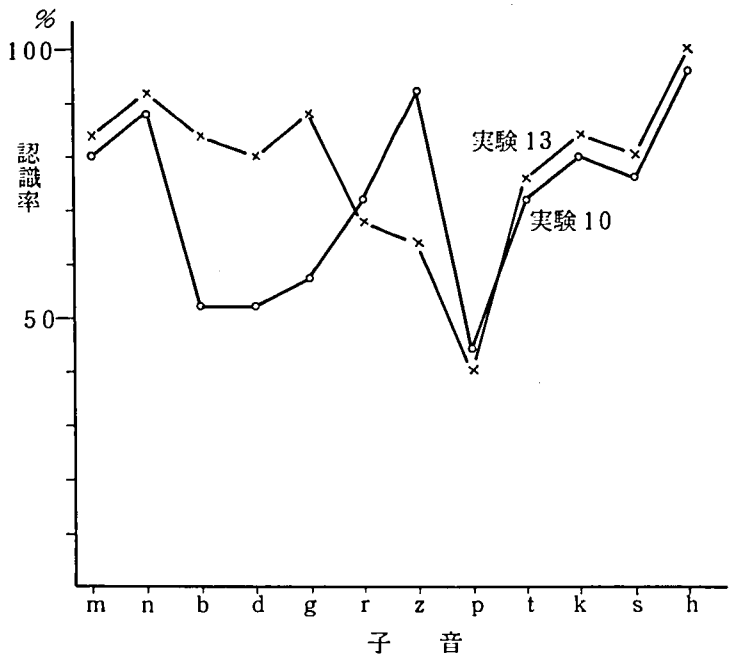


図 3.18　子音別の認識率
（学習サンプルの組数と認識率の関係）

以上のように，学習サンプルの組数をふやして，パターンの変動に対する平均化のされた標準パターンを作った方が高い認識率が得られる。なお，実験13では入力音声の母音部分はわかっているものとしてあるが，実験10の結果を見てもわかるように母音の認識は安定しており，ほぼ確実に認識できると考えられるから実験10と13を直接比較してもかまわないといえよう。

(4)　基準サンプル

基準サンプルの選び方が認識にどのような影響を及ぼすかを調べる。まず基準サンプルの選び方が実際に認識結果に影響するかどうかを調べるため実験2と3を比較する。実験2では未知サンプルを基準サンプルとして用いており，実験3では学習サンプルのうち1組を基準サンプルとして用いている。図3.19に順位と認識率のグラフを示すが実験2の結果の方が良いことがわかる。これは次のように説明することができる。実験2では未知サンプルを基準サンプルとして用いたため，標準パターンの時間構造が未知サンプルに似てくるものと思われる。そのため，未知サンプルに関する情報を持たない標準パターンを使った実験3より認識率が良いのであろう。

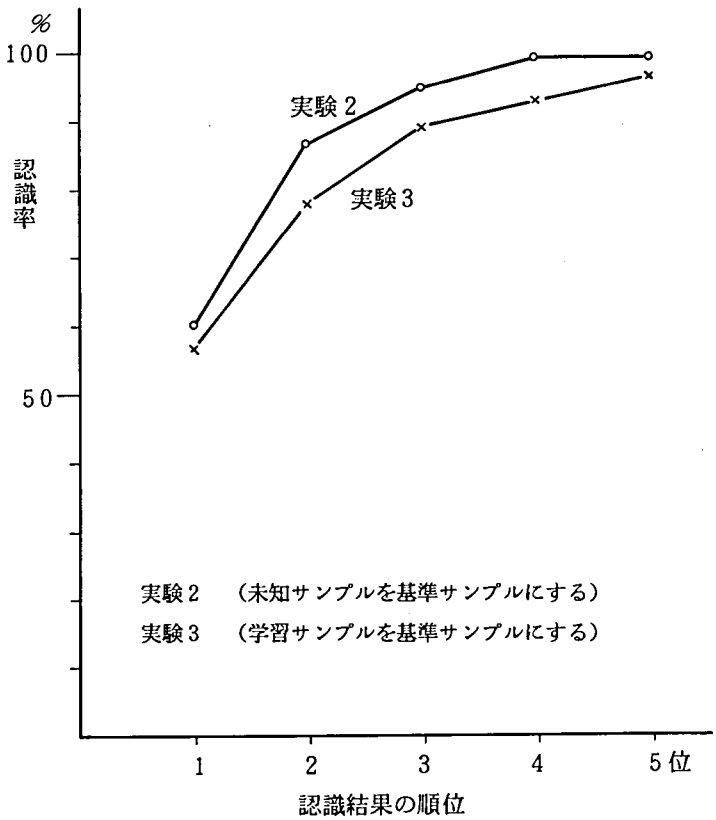


図3.19　順位と認識率のグラフ
（基準サンプルの選び方と認識率の関係）

上で述べたように，基準サンプルの選び方は認識結果にかなり大きな影響を及ぼすことがわかる。したがって，標準パターンがすべての未知サンプルに対する適応性を持つためには，1組のサンプルを基準サンプルとするより，何組かのサンプルを平均して基準サンプルとした方が良いであろう。このような観点から作成した標準パターンE_2を標準パターンE_1と比較するため実験13と14の結果を図3.20（順位と認識率のグラフ）および図3.21（子音別の認識率）に示す。これらの図から実験14の方が良い結果であることがわかる。したがって，予想通り標準パターンE_2の方が有効である。ただし，その差はわずかであり，今後データを増やす等をして確認する必要があろう。

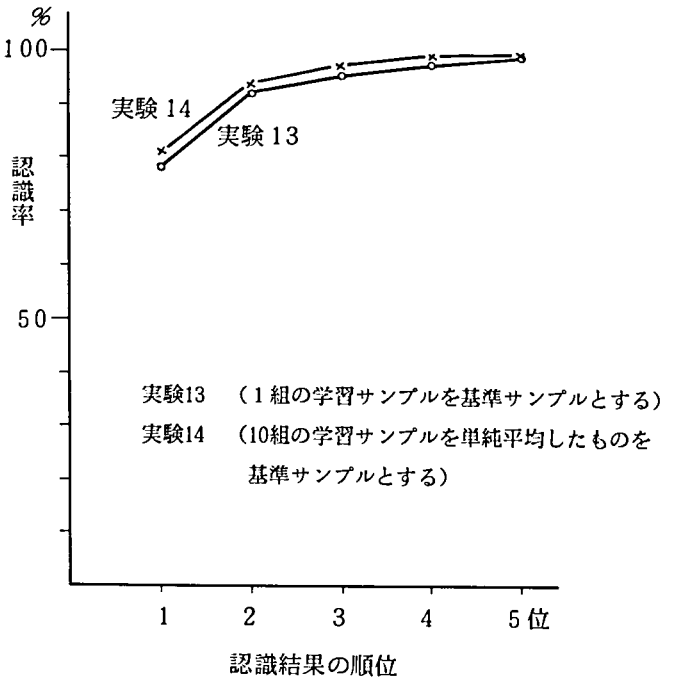


図 3.20　順位と認識率のグラフ
（基準サンプルの選び方と認識率の関係）

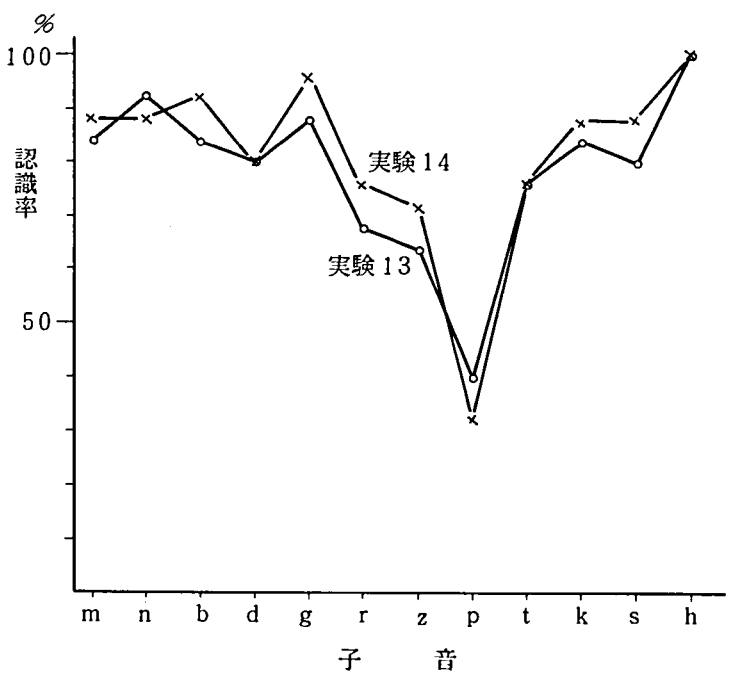


図 3.21　子音別の認識率
（基準サンプルの選び方と認識率の関係）

(5) フレームの場所

ＶＣＶ音節の認識には，ＶＣＶ音節のどの部分が大きく影響しているかを知るため標準パターンC_1を使った実験７と標準パターンC_2を用いた実験８を比較する。

図3.22に順位と認識率のグラフを示す。標準パターンC_1の方がC_2よりフレーム数が少ない（C_1のフレーム数はもとのＶＣＶ音節の⅓，C_2は½）にもかかわらず実験７の結果が良いことがわかる。したがって，ＶＣＶ音節の認識には中央の子音部分のみが重要なのではなく，母音から子音への過渡部分を含んだ広い範囲にわたる情報が必要なことがわかる。

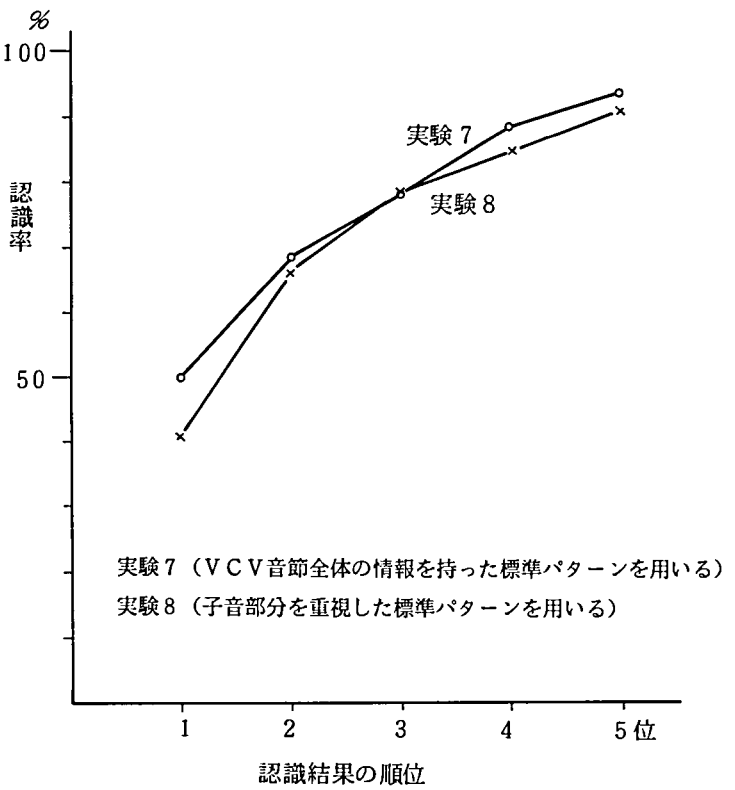


図 3.22　順位と認識率のグラフ
（フレームの位置と認識率の関係）

(6) 学習による標準パターンの改良

標準パターンE_3はE_2を基準サンプルとして作りなおしたものである。この方法を拡張すれば，出来上った標準パターンを基準サンプルとして，改めて標準サンプルを作りなおすことをくりかえすという，一種の学習により最適な標準パターンが作成できると考えられる。この考え方が正しいかどうかを調べるため，標準パターンE_2を用いた実験15とE_3を用いた実験20の比較をおこなう。図3.23に順位と認識率のグラフを，図3.24に子音別の認識率を示す。これらの結

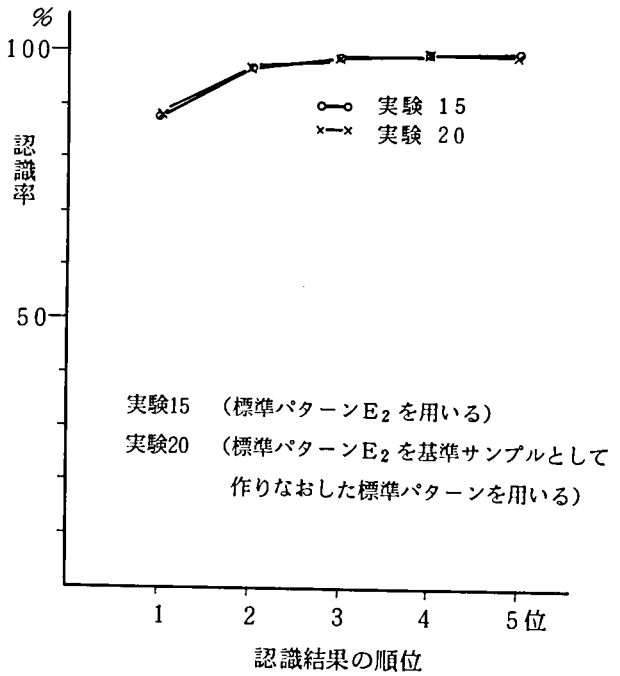


図 3.23 順位と認識率のグラフ
（学習による標準パターンの改良の効果）

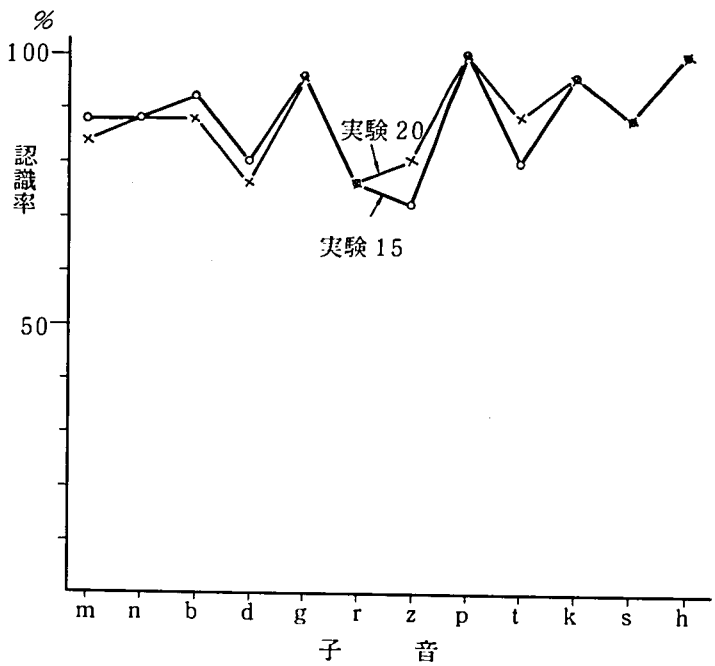


図 3.24 子音別の認識率
（学習による標準パターンの改良の効果）

果をみると両方の結果にほとんど差がないことがわかる。したがって1回の学習で作られた標準パターンE_2は十分に収束しており，何度も学習をくりかえしてもあまり標準パターンの改良はされないといえる。

以上の検討をまとめると標準パターン作成法について次の結論がいえる。

(1) 音韻，ＶＣ，ＣＶ音節，ＶＣＶ音節の3種類の音声単位の中で，認識に関してはＶＣＶ音節が最も有利である。

(2) 標準パターンのフレーム数は多い方が，すなわちスペクトルパターンの時間推移に関する情報を多く含む方が有利である。

(3) 学習サンプルの組数は多い方が有利である。

(4) 基準サンプルは特定のサンプルを用いるのは望ましくない。むしろ，いくつかのサンプルを単純平均したものを用いる方が有利である。

(5) 標準パターンは，中央の子音部分に関する情報のみをくわしく持つより，ＶＣＶ音節全体にわたる情報を持った方が有利である。

(6) 学習による標準パターンを作りなおしても大きな効果はない。

3.6.2 認識法の検討

(1) パワー情報の使用

入力音声のパワー情報を利用する効果を知るため実験14と15の結果を比較する。図3.25に順位と認識率のグラフを，図3.26に子音別の認識率を示す。パワーの情報を使うことにより認識率は6.7％上昇している。図3.26を見ると，これは/p, t, k/の認識率が上ったためである。/p, t, k/の認識率が上ったのは，実験14では/p, t, k/→/b, d, g/なるconfusionが多かったのがパワーの情報を使うことにより，有声，無声の判別が容易にでき，このようなconfusionがほとんどなくなったことが原因である。(付録1参照) したがって，スペクトルパターンのマッチングを基本とした認識法にパワーのような異なった情報にもとづいた認識法を加味することにより結果がかなり改善されることがわかる。ただし，図3.10の判定理論は2名の発声者から集めたデータにもとづいて作ったものであるから今後データを増やして検討を加える必要がある。

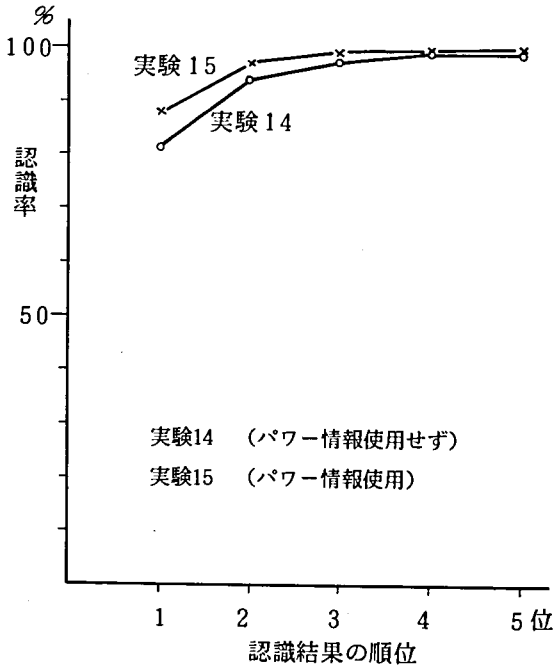


図 3.25　順位と認識率のグラフ
(パワー情報を使用する効果)

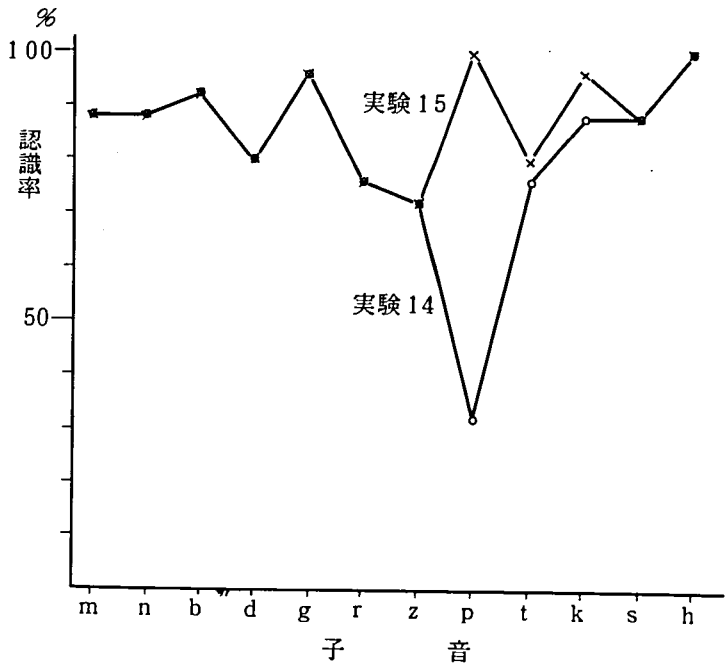


図 3.26　子音別の認識率
(パワー情報を使用する効果)

(2) 重みづけ

4.2.2で述べたように重みづけの効果を調べるため，実験14と17の比較をおこなう。順位と認識率のグラフを図3.27に，子音別の認識率を図3.28に示す。これらの図から，式（3.28）の

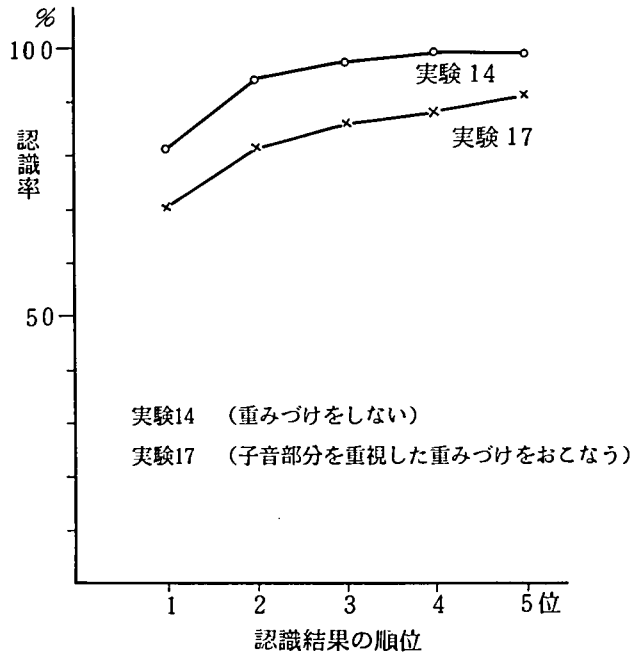


図 3.27 順位と認識率のグラフ（重みづけの効果）

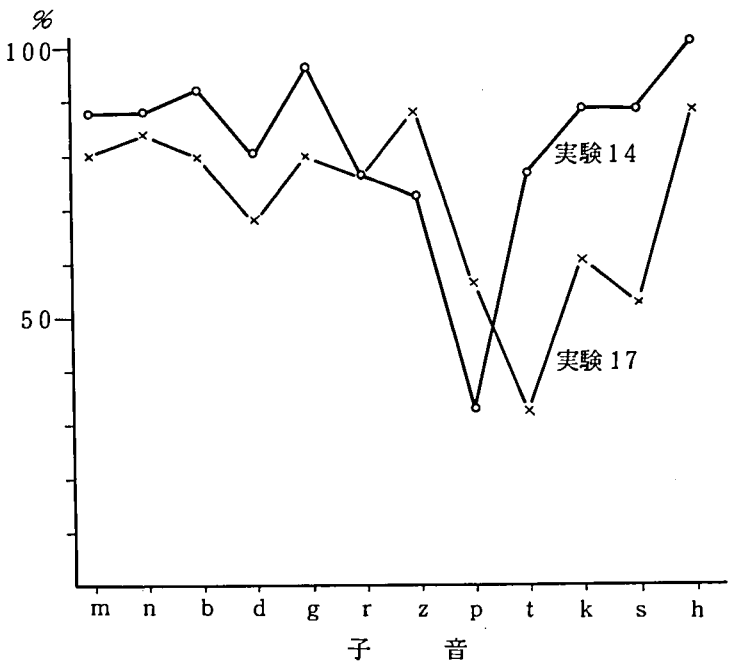


図 3.28 子音別の認識率（重みづけの効果）

重みづけが認識にかなり悪影響を及ぼしていることがわかる。このことはＶＣＶ音節の認識には中央の子音部分のみが関係しているのではなく，両端の母音をも含めた広い範囲にわたるスペクトルパターンの変化が関係していることを示している。これは標準パターンの作成法の検討の時に得られた結論と同じである。

(3)　継続時間の制限

4.2.3で述べたように，標準パターンの１フレームに対応づけられる未知サンプルのフレーム数に制限をつけることによる効果を調べる。実験14と実験18の結果を図3.29（順位と認識率のグラフ）及び図3.30（子音別の認識率）に示す。図3.29を見ると，継続時間の制限をつけることが有効とはいえず，かえって認識率を下げていることがわかる。ただし，図3.30をを見ると実験14では子音ごとの認識率にばらつきが多いのに対し，実験18では割合，ばらつきが平均化されていることがわかる。ただし，このような傾向が一般的なものかどうかについてはより多くのデータを集めてたしかめる必要があろう。

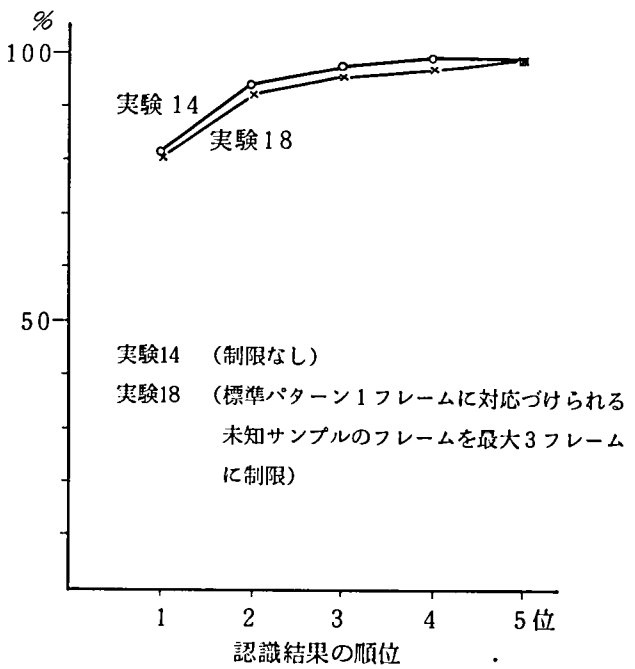


図3.29　順位と認識率のグラフ
（継続時間の制限の効果）

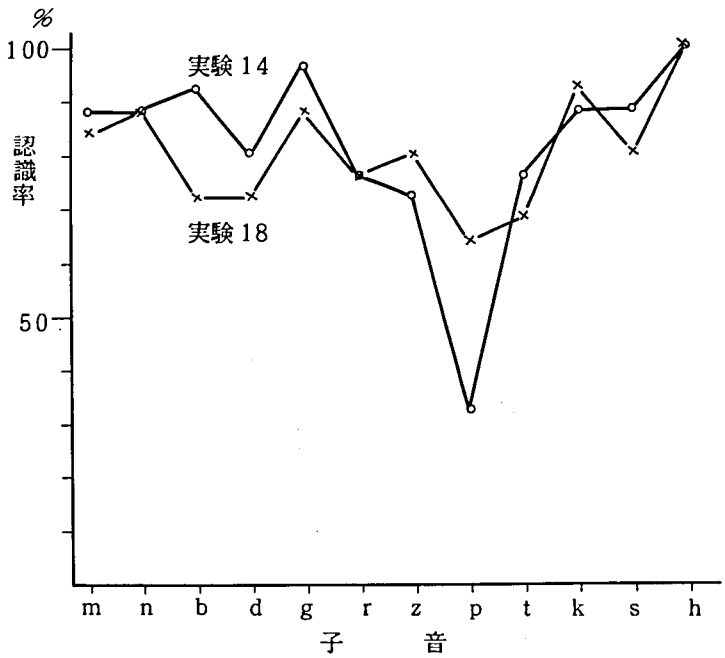


図 3.30　子音別の認識率
（継続時間の制限の効果）

(4)　cut off frequency

cut off frequency を 6 kHz にすることの効果を調べるため，実験15と19の順位と認識率のグラフを図 3.31に示す。ただし実験15については有声子音のみについての結果を示してある。また，子音別の認識率を図 3.32に示す。実験19の認識対象は有声子音のみよりなる 175 種類のＶＣＶ音節であるが，実験15の confusion matrix（付録 1 参照）を見るとわかるように有声子音と無声子音の間の confusion はほとんどないため，実験 15 の有声子音に関する結果を実験19の結果と比較してもさしつかえない。図 3.32で注目されるのは実験15で最も認識率の低い / z / が実験19では 100 %になっていることである。/ z / は 3.2 kHz より高域に多くの成分を持っているため，cut off frequency を上げた効果があらわれたものと考えられる。他の有声子音は高域にそれ程成分を持っていないため，改善はあまり認められない。

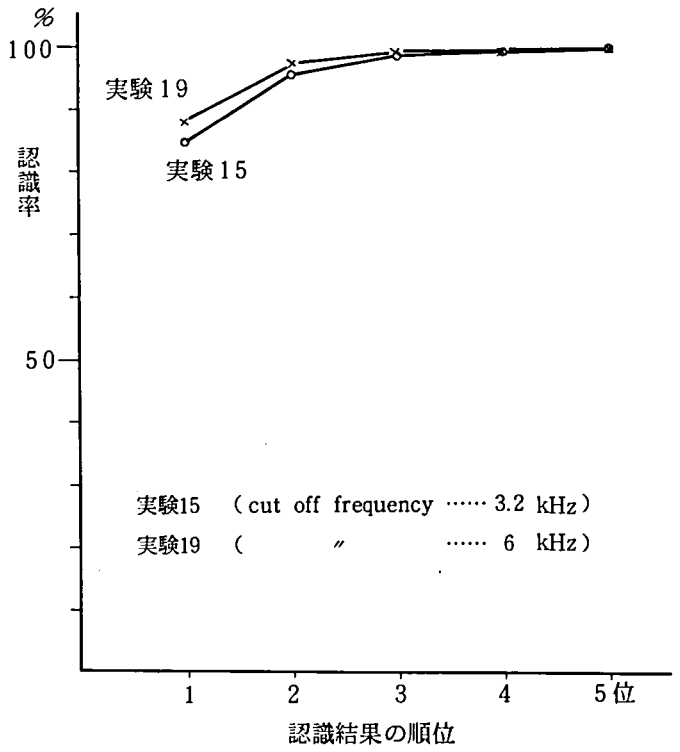


図3.31 順位と認識率のグラフ
(cut off frequencyの影響)

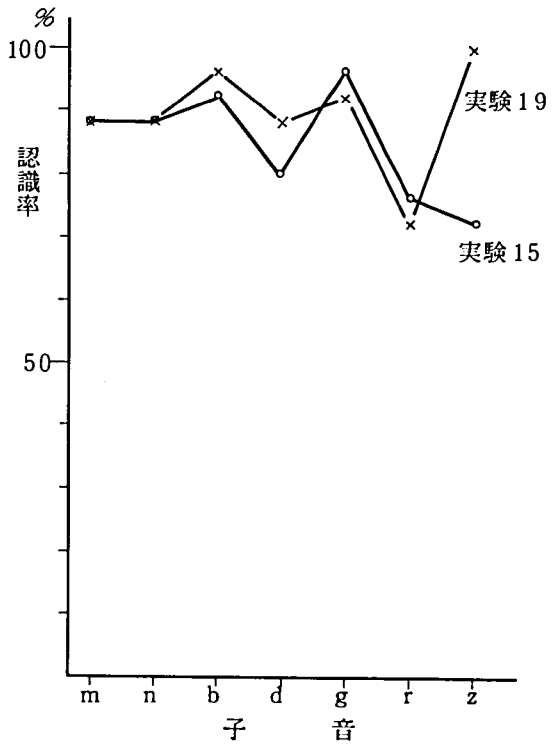


図3.32 子音別の認識率
(cut off frequencyの影響)

(5) 学習サンプルの認識

未知サンプルと学習サンプルの認識結果を比較検討する。まず実験 2, 4, 5, 6.の結果を順位と認識率の関係について図 3.33に示す。実験 2 は未知サンプルの認識実験，実験 4 は学習サンプルの認識実験である。実験 5, 6 で用いた標準パターンは実験 2, 3 で用いた標準パターンを平均して作ったものである。したがって実験 5, 6 は学習サンプルの組数をふやして作った標準パターンを用いた学習サンプルの認識実験とみなすことができる。

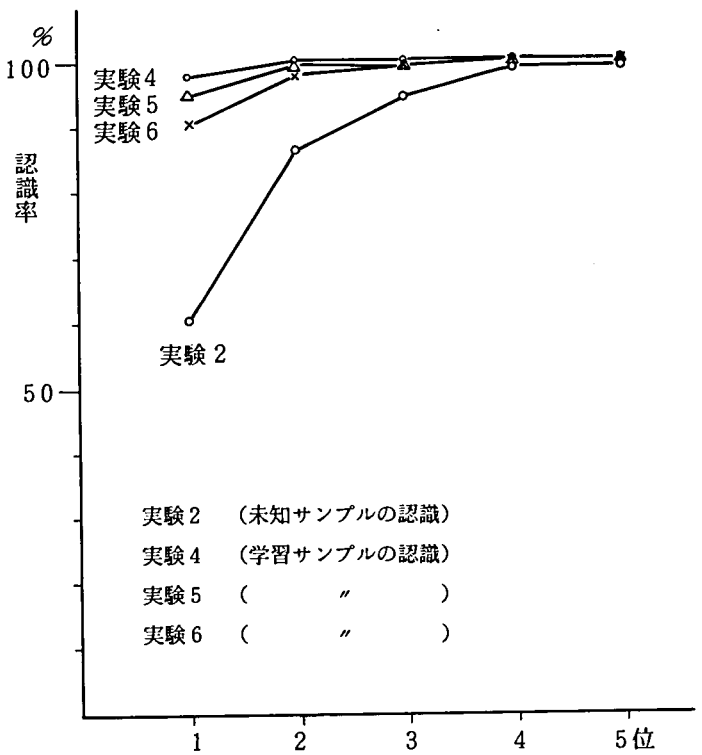


図 3.33 順位と認識率のグラフ
(学習サンプルと未知サンプルの認識結果の比較)

図 3.33より次のことがいえる。

- 学習サンプルと未知サンプルの認識率にはかなりの差がある。これは，標準パターンが不完全なものであって，未知サンプルに対する適応性をもっていないことを示している。
- 学習サンプルの組数を増やすと，学習サンプルの認識率は低下する。すなわち，学習サンプルの組数を増やすと，パターンの変動がより平均化された標準パターンが作られるため，学習サンプルと未知サンプルの認識率の差が少なくなるものと思われる。上に

述べたことをたしかめるため，実験15と16を比較する。図 3.34に順位と認識率のグラフを，図 3.35に子音別の認識率を示す。図 3.34，3.35を見ると実験16の方が良く，子音別

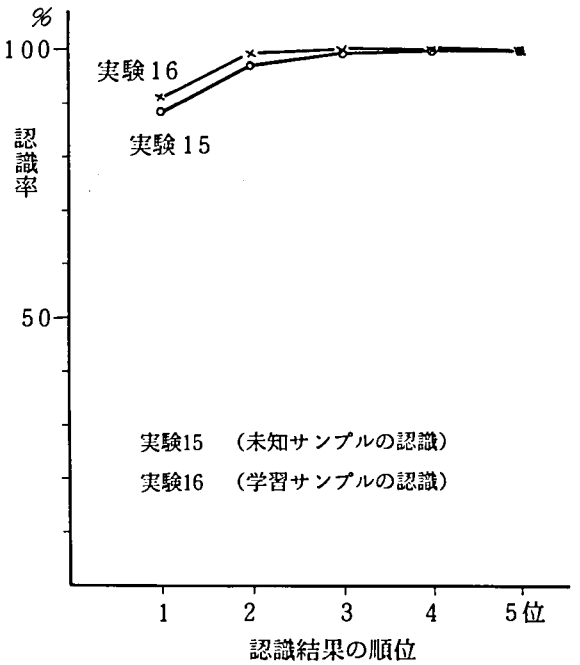


図 3.34　順位と認識率のグラフ
(学習サンプルと未知サンプの認識結果の比較)

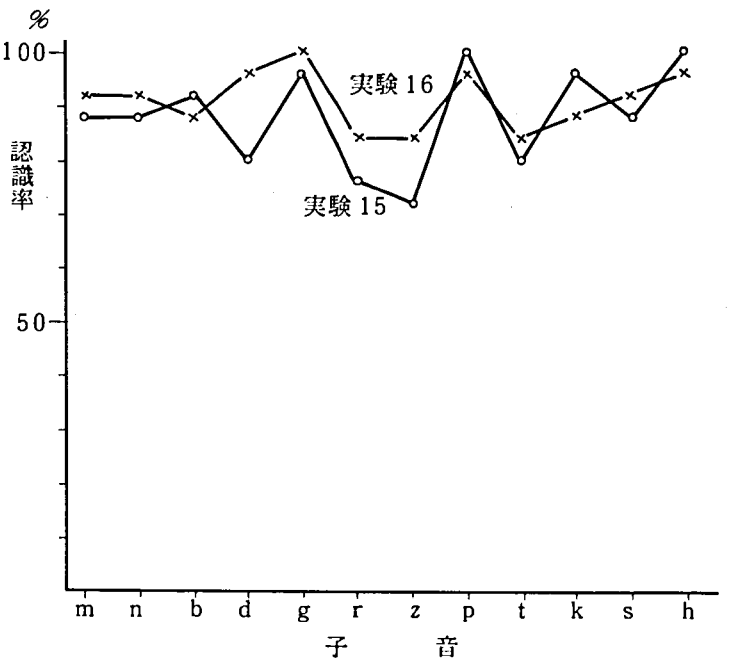


図 3.35　子音別の認識率
(学習サンプルと未知サンプルの認識結果の比較)

の認識率も安定しているがその差はわずかである。これは標準パターンが変動に対し十分平均化されたものであることを示している。また学習サンプルと未知サンプルの認識率にあまり差がないことから，これ以上学習サンプルの組数を増やしても認識率はそれほど向上しないものと思われる。

以上の検討をまとめると認識法については次の結論がいえる。

(1) パワーの情報を用いることは認識に有利である。

(2) ＶＣＶ音節の認識の際，中央の子音部分のみ重視したパターンマッチングは好ましくない。ＶＣＶ音節全体にわたるマッチングが必要である。

(3) パターンマッチングの際，標準パターンの１フレームに対応づける未知サンプルのフレーム数に制限をつけることはあまり効果がない。

(4) cut off frequency をあげると，高域に成分を持つ子音の認識率はあがる。低域に成分が集中している子音はあまり改善されない。

(5) 標準パターン作成のための学習サンプルの組数をふやすと，学習サンプルの認識は下がるが，未知サンプルの認識率は上がる。すなわち，標準パターンの未知サンプルに対する適応性が増す。学習サンプルを10組用意して作った標準パターンでは，学習サンプルと未知サンプルの認識率の差があまりなくなるから，学習サンプルの組数は 10 組程度で十分である。

3.7 あとがき

ＶＣＶ音節を音声単位として連続音声を認識するシステムを構成するための第一歩として，単独に発声された 300 種類のＶＣＶ音節の認識実験をおこなった。まず，いくつかの学習サンプルを時間軸の正規化をおこなった後平均して標準パターンを作成する方法を提案した。他の音声単位との比較をおこなうため，音韻やＶＣ，ＣＶ音節を音声単位とした標準パターンを作り比較実験をおこなった。その結果，認識に関してはＶＣＶ音節を音声単位にとるのが最も有効であることがわかった。また，学習サンプルの組数，認識の際の制限等を種々かえて実験をおこなった結果，10組の学習サンプルより作った標準パターンを用いて，パワー情報を補助情報として認識すると学習サンプルで 91.0％，未知サンプルで 88.0％の認識率が得られた。ＶＣＶ音節において両側の母音はほぼ確実に認識できるから，ＶＣＶ音節の認識率はそのまま中央の子音の認識率と考えてよい。そのような観点からすると８８.０％という認識率はかなり高いもの

である。

以上の検討により，ＶＣＶ音節を単位とした音声認識法の見通しが得られた。また，本章で提案した，複数個の時系列パターンを時間軸の正規化を行った後平均して標準パターンを作成する方法は，その後他研究機関でも用いられるようになっており，[117] 現在では一般的な手法として認められている。

第4章　VCV音節を単位とした単語音声の認識

4.1　はしがき

第3章では，VCV音節の認識の問題を取り扱った。本章では，次の段階として，VCV音節を単位とした単語音声の認識について述べる。いきなり最終的な目標である会話音声の認識を目ざすのではなく，途中段階として単語音声の認識を試みるのは次のような理由による。

(1)　従来，単語音声認識においては，セグメンテーション→セグメントの認識→単語の認識，という手順をふんで認識する方法は，単語単位で標準パターンとのマッチングを行う手法に比較してあまり高い認識率を得ることは出来なかった。これに対し，VCV音節を単位にとった場合，単語を単位とした方法に匹敵するだけの認識率が得られたとすれば，それは，VCV音節を単位にとることの有効性を示すことになる。

(2)　連続単語音声，会話音声などのより高度な対象を取り扱う際の基礎データになる。

以上のような観点から，本章では，日本語より選んだ90種類の単語を認識対象とし，VCV音節を単位とした単語音声認識について述べる。

4.2　認識対象

日本語から選んだ90種類の単語を認識対象とする。それらの単語のうちわけは次の通りである。

VCV型単語 ……30種類

VCVCV型単語 ……30種類

VCVCVCV型単語……30種類

単語のリストを表4.1に示す。この90種類の単語を8名の男性発声者（T，Ki，F，N，I，S，H，Ko）が各5回発声した計3,600サンプルを用意する。これらのうち，語尾の母音が無声化したサンプルを除いた3,580サンプルを認識実験に用いた。認識対象の単語を構成している音韻は次に示す母音5種類および子音12種類である。

表4.1 認 識 対 象 の 単 語

VCV型　　uta（歌），oba（おば），eki（駅），imu（医務），ato（後），oni（鬼），iro（色），one（尾根），aze（あぜ），isu（いす），ichi（一），oku（奥），ura（裏），ase（汗），ima（居間），esa（えさ），igo（囲碁），ada（仇），ushi（牛），oto（音），eri（えり），ebi（海老），ehu（F），ume（梅），uha（右派），azi（味），edo（江戸），ike（池），ono（オノ），igi（異議）

VCVCV型　　omote（表），ihuku（衣　服），akebi（あけび），abura（油），umibe（海辺），otsuge（お　告），odosu（威す），onushi（お　主），agari（上　り），aniki（兄　貴），akago（赤子），eboshi（えぼし），umare（生れ），ichiza（一　座），ikusa（戦），ahiru（アヒル），otaku（お宅），enogu（絵　具），udegi（腕木），asobi（遊　び），irori（いろり），otoko（男），ibara（茨），opera（オペラ），umezu（梅酢），adana（あだ名），asemo（あせも），ohana（お　花），eziki（餌食），uneri（うねり）

VCVCVCV型　amehuri（雨降り），azakeri（喚り），anaguma（穴熊），ibaragi（茨木），inemuri（居眠），udegumi（腕 組 み），umeboshi（梅 干 し），esupuri（エスプリ），ogakuzu（おがくず），odoriko（踊 り 子），ohitashi（おひたし），ezomatsu（え ぞ 松），ibukuro（胃　袋），azemichi（あ ぜ 道），ichiziku（いちじく），ehagaki（絵 葉 書），ebigani（えびがに），usemono（う せ 物），otedama（お 手 玉），otohime（乙　姫），akagire（あかぎれ），asakaze（朝　風），agezoko（あ げ 底），agohige（あごひげ），arupaka（アルパカ），inutade（いぬたで），usotsuki（うそつき），abekobe（あべこべ），unohana（う の 花），uketsuke（受　付）

母音……／a／，／i／，／u／，／e／，／o／

子音……／m／，／n／，／b／，／d／，／g／，／r／，／z／，／p／，／t／，／k／，／s／，／h／

したがって，標準パターンとして用意するVCV音節は300種類である。これらの標準パターンは各発声者ごとに用意したものを用いる。標準パターンの作成法については4.4.2.1で述べる。

4.3 入力処理

入力音声は第2章で述べた音響処理系により処理されて，特徴パラメータの時系列に変換される。すなわち，まず3.2kHz低域通過フィルタを通ったあと標本化周波数8kHzのAD変換器でディジタル音声に変換される。

次に，ディジタル音声を15msecごとのフレームに分け，各フレームのパワーを計算する。パワーがあらかじめ与えられた閾値より始めて大きくなったフレームを音声区間の始端とする。閾値以下のフレームが30フレーム（450msec）以上続いた時は，最初に閾値以下になったフレームを音声区間の終端とする。音声区間内でパワーが閾値以下になる区間は無音区間と呼ぶことにする。無音区間は無声子音の前に生じる場合と，母音が無声化したために生じる場合とがある。前者の場合を休止区間，後者の場合を無声化区間と呼ぶことにする。

以上の手続きにより音声区間が決定される。音声区間においては，15msecのフレームごとに，音声の特徴量として音声波形の自己相関係数

$$\boldsymbol{\rho} = (\rho_0, \rho_1 \cdots\cdots, \rho_p) \qquad (4.1)$$

および音声パワーv_0を求める。

したがって，各単語はρの時系列

$$\boldsymbol{P} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \cdots\cdots, \boldsymbol{\rho}_N) \qquad (4.2)$$

で表わされる。ただし

$$\boldsymbol{\rho}_i = (\rho_{i0}, \rho_{i1}, \cdots\cdots, \rho_{ip}) \qquad (4.3)$$

は第 i フレームの自己相関関数である。また，音声のパワー値の系列

$$P = (v_{10},\ v_{20},\ \cdots\cdots,\ v_{N0}) \qquad (4.4)$$

をパワー系列と呼ぶことにする。認識系へ送られる情報は，式（4.2），および式（4.4）である。

4.4 認識系の構成[23][118]

認識系は図4.1に示したように，大きく分けるとセグメンテーション部と認識部から構成される。以下，各部について説明する。

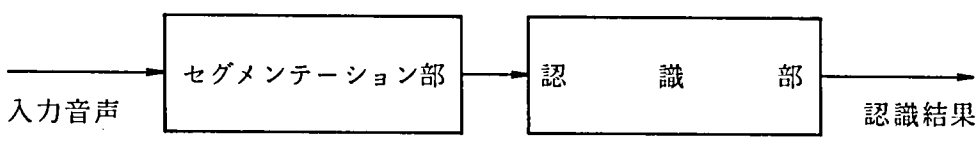


図 4.1　単語音声認識系の構成

4.4.1 セグメンテーション部

4.4.1.1 セグメンテーションの方法

セグメンテーションは困難な問題であり，現在まで種々の方法が試みられてきたが，まだその方法が確立していないのが現状である。これは，本来連続的な変化をしている音声を不連続な単位に切ろうとするところに原因がある。したがって，万能なセグメンテーションの方法を考えるより，音声単位のとり方に依存した方法を採用する方が有利である。また，当然セグメンテーションの段階で誤りが生じることを考慮し，できれば認識の段階で誤りの回復ができるような方法を考える必要がある。以上の点を考慮し，ここでは次の方針に従ったセグメンテーションの方法を採用する。

(1) VCV音節を単位としたセグメンテーションをするには母音部分を検出すればよいが，一般に母音部分はパワーが大きく，しかもこの性質は個人差，発声速度に影響されにくい。そこで，パワーの変化を主な情報としてセグメンテーションを行う。また，継続時間は個人差，発

声速度による変動が大きいが，比較的簡単に使える情報であるため，これを副次的な情報として用いることにする。

(2) セグメンテーションの際の誤りとしては，セグメント境界の検出ができなかった誤り(脱落誤り）と，実際はセグメントの境界でないところを境界として検出した誤り（挿入誤り）の2種類がある。高いセグメント化率を得るという観点からすると両者の誤りがバランスするような方法をとるべきであるが，ここでは認識系全体の性能，構成の容易さという観点から，挿入誤りをある程度認める代りに脱落誤りをできるだけ少なくするという方法をとった。具体的には，セグメントの境界を一意に決めるのが困難な場合にはいくつか候補点をあげ，認識の段階で候補点のうちで最適のものを選択するという方法をとった。このような方法をとることにより，セグメンテーション部，認識部共に構成法が容易になったばかりでなく，認識部にセグメンテーション誤りをある程度回復できる機能をもたせることができた。

セグメンテーション部の構成を図4.2に示す。具体的なセグメンテーションの手順は次の通りである。

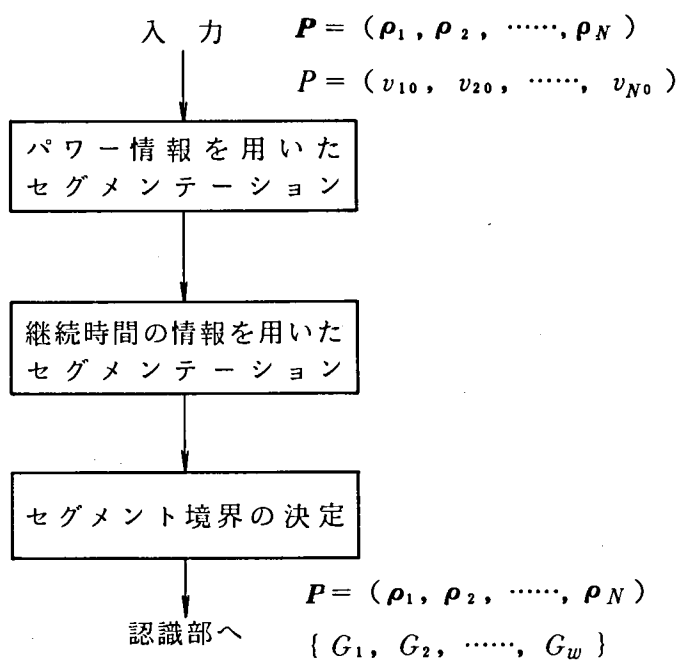


図4.2　セグメンテーション部の構成

STEP1(パワー情報を用いたセグメンテーション)

(1) 入力音声のパワーは式（4.4）で与えられるものとする。式（4.4）の系列から微少な凸凹を除くため，次のような平滑化を加える。すなわち，

$$\tilde{v}_i = (\sqrt{v_{i0}} + \sqrt{v_{i+1,0}}) / 2 \qquad (4.5)$$

$(i = 1, 2, \cdots\cdots, N-1)$

とし，

$$\tilde{P} = (\tilde{v}_1, \tilde{v}_2, \cdots\cdots, \tilde{v}_{N-1}) \qquad (4.6)$$

が新しくパワーの情報を与えるものとする。

(2) あらかじめ閾値P_Tを与えておき，$\tilde{v}_j < P_T$なる点を除くことにより，式（4.6）を次に示すようないくつかの小系列に分割する。

$$\tilde{P}_1 = (\tilde{v}_{n_1}, \tilde{v}_{n_1+1}, \cdots\cdots, \tilde{v}_{n_2}),\ \tilde{P}_2 = (\tilde{v}_{n_3}, \tilde{v}_{n_3+1}, \cdots\cdots, \tilde{v}_{n_4})$$

$$\cdots\cdots\ \tilde{P}_t = (\tilde{v}_{n_{2t-1}}, \cdots\cdots, \tilde{v}_{n_{2t}}) \qquad (4.7)$$

$(1 \leqq n_1 < n_2 < \cdots\cdots < n_{2t} \leqq N-1)$

(3) 式（4.7）より任意の$\tilde{P}_i$を選び出す。$\tilde{P}_i$の中の$\tilde{v}_j$の系列から極小点を検出し，あわせて前後の極大値との差を計算する。ある極小点における前後の極大値との差をdv_1, dv_2とする。あらかじめ閾値Q_Tを与えておき，

$$dv_1 > Q_T \quad \text{かつ} \quad dv_2 > Q_T \qquad (4.8)$$

であれば，その極小点を$\tilde{P}_i$から除去する。この操作をすべての$\tilde{P}_i$に加えることによって，$\tilde{P}$は更に短い系列，

$$\tilde{\tilde{P}}_1 = (\tilde{v}_{m_1}, \tilde{v}_{m_1+1}, \cdots\cdots, \tilde{v}_{m_2}),\ \tilde{\tilde{P}}_2 = (\tilde{v}_{m_2+1}, \cdots\cdots, \tilde{v}_{m_3})$$

$$\cdots\cdots\ \tilde{\tilde{P}}_u = (\tilde{v}_{m_{2u-1}}, \cdots\cdots, \tilde{v}_{m_{2u}}) \qquad (4.9)$$

$(1 \leqq m_1 \leqq m_2 < \cdots\cdots < m_{2u} \leqq N-1)$

へ分割される。以上の手順を図4.3に示す。

STEP 2（継続時間の情報を使ったセグメンテーション）

鼻音等ではパワーが小さくならない場合が多いので，以上の操作で分割された各区間内には

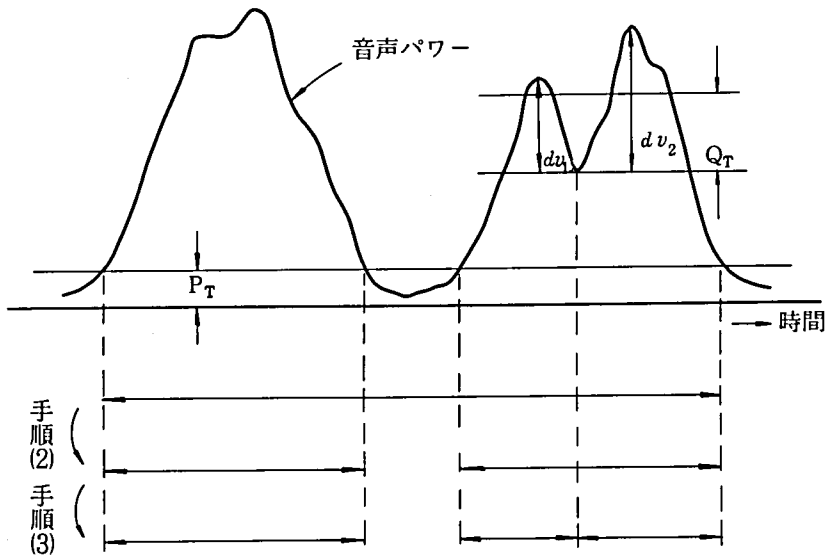


図 4.3　パワー情報を用いたセグメンテーション
（step 1）

手順(2)　無音区間検出によるセグメンテーション
手順(3)　パワー極小点検出によるセグメンテーション

1個ないし数個の母音が含まれている。次に，継続時間の情報を使って，区間内に含まれる母音数を決定する。

式（4.9）の任意の区間$\tilde{\tilde{P}}_i$のフレーム数をl_iとする。閾値の集合

$$\{ S_k{}^m \}, \quad \{ S_k{}^M \} \qquad (k = 1, 2, \cdots\cdots) \tag{4.10}$$

を用意し，

$$S_k{}^m \leqq l_i \leqq S_k{}^M \tag{4.11}$$

ならば，$\tilde{\tilde{P}}_i$中にはk個の母音が含まれている可能性があると判定する。

STEP 3（セグメント境界の決定）

l_iなるフレーム数をもつ区間内にk個の母音が含まれていると判定されると，区間の始端から，

$$\frac{2j-1}{2k} l_i \quad (j=1, 2, \cdots\cdots, k) \qquad (4.12)$$

番目のフレームをセグメントの境界とする。このようにして，$\tilde{\tilde{P}}_i$中の一組のセグメント境界が決定する。継続時間に関する閾値の決め方によっては，$\tilde{\tilde{P}}_i$中に含まれると判定される母音の数は一種類とは限らない。したがって，$\tilde{\tilde{P}}_i$にはセグメント境界の組の集合が対応づけられる。これを，

$$\{ G_1^i, G_2^i, \cdots\cdots, G_{i_n}^i \} \qquad (4.13)$$

であらわす。ただし，各G_j^iの要素はセグメントの境界のフレーム番号である。(各フレームは音声区間の先頭から順に番号がつけられているものとする)

以上の操作を各区間に適用する。各$\tilde{\tilde{P}}_i$に対応づけられた集合の要素を1つずつとりだしてできる。

$$G = G_{j_1}^1 \cup G_{j_2}^2 \cup \cdots\cdots \cup G_{ju}^u \qquad (4.14)$$

は，入力音声の一組のセグメント境界を表わしている。このようにしてできる，全ての異なったGを

$$\{ G_1, G_2, \cdots\cdots, G_w \} \qquad (4.15)$$

とするとこれらがセグメントの境界の候補になる。STEP 2，STEP 3の手順を図4.4に示す。

4.4.1.2 無声化の対策

実験に用いたサンプル中には，語中で無声化しやすいもの（たとえば／ohitashi／等）があるのでその対策を考えておく必要がある。

一般に，無声化した場合に生じる無声化区間は休止区間に比べ長いので，次のような簡単な方法で無声化と休止の判定をする。

無声化区間のフレーム数をlとする。2つの閾値S_1，S_2を定めておき，

$$\left.\begin{array}{l} l \leqq S_1 \text{の場合}\cdots\cdots\text{休止区間} \\ l \geqq S_2 \text{の場合}\cdots\cdots\text{無声化区間} \end{array}\right\} \qquad (4.16)$$

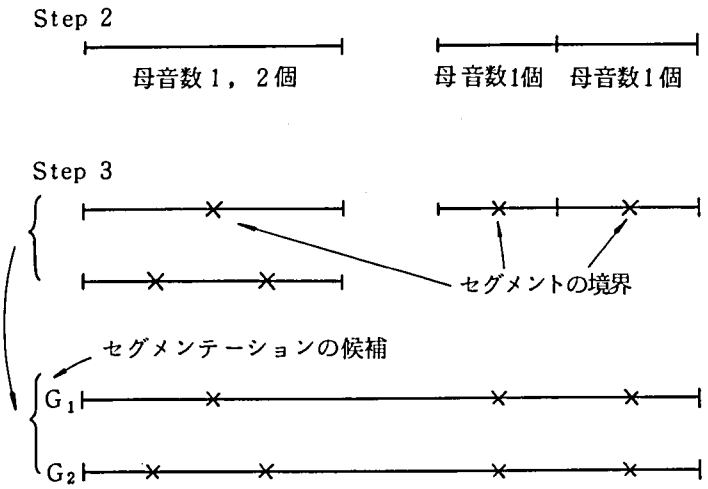


図 4.4　セグメンテーションの手順
（step 2，step 3）

と判定する。無声化区間と判定された場合は，区間の中央に新しくセグメントの境界をもうける。したがって，無音区間にも式（4.13）に対応するセグメントの境界の組が対応づけられるわけで，これを，

$$\{ \widetilde{G}_1^i,\ \widetilde{G}_2^i,\ \cdots\cdots,\ \widetilde{G}_{i_n}^i \} \qquad (4.17)$$

で表す。ただし，無声化区間か休止区間かの２つの可能性しかないため，i はたかだか２である。また，休止と判定された場合は，セグメントの境界が存在しないので，このとき $\widetilde{G}_j^i$ は空集合となる。

以後，セグメント境界の組の集合，式（4.15）は無音区間のセグメント境界も含めて考えたものとする。したがって，セグメンテーション部から認識部へ送られるのは自己相関係数の時系列（4.2），パワーの系列（4.4），セグメンテーションの候補（4.15）である。

4.4.2 認識部

4.4.2.1 標準パターン作成法

標準パターンとして300種類のVCV音節を各発声者ごとに用意した。標準パターンを作成する際，次の方針に従うものとした。

(1) VCV音節を音声単位にとった利点を生かすため，母音と子音の間の過渡部分を含んだ標準パターンであることが望ましい。

(2) 種々の入力に対する適応性を持つため，パターンのばらつきに対し十分平均化された標準パターンであることが望ましい。

上の方針に従った種々の標準パターン作成法については第3章でくわしく述べた。特に(2)については，いくつかの学習サンプルを時間軸の非線形な伸縮の補正を行った後，平均することにより標準パターンを作成する方法を提案し，その有効性を示した。そこで，ここでは，第3章で示した標準パターンのうち，認識率が高く，かつ作成法の比較的容易な標準パターン E_2 を用いることにする。

次に作成法を簡単に述べる。

(1) 各VCV音節ごとに10組の学習サンプルを用意する。特定のVCV音節に注目し，その学習サンプルを

$$\boldsymbol{P}_j = (\boldsymbol{\rho}_1^j, \boldsymbol{\rho}_2^j, \cdots\cdots, \boldsymbol{\rho}_{M_j}^j) \qquad (4.18)$$

$$(j = 1, 2, \cdots\cdots, 10)$$

とする。ただし，M_j は j 番目の学習サンプルのフレーム数であり，$\boldsymbol{\rho}_i^j$ は j 番目の学習サンプルの第 i フレームの自己相関係数である。

(2) 学習サンプルを平均して基準サンプル

$$\boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots\cdots, \boldsymbol{q}_M) \qquad (4.19)$$

を作る。$\boldsymbol{q}_i$ は次式で与えられる。(図4.5)

$$\boldsymbol{q}_i = \frac{1}{10} \sum_{j=1}^{10} \boldsymbol{\rho}_i^j \qquad (4.20)$$

$$(i = 1, 2, \cdots\cdots, M : M = \min(M_1, M_2, \cdots\cdots, M_{10}))$$

(3) 次に，$\boldsymbol{Q}$ を基準として各学習サンプルの時間軸の正規化を行う。そのためにまず，$\boldsymbol{\rho}_i^j$ と

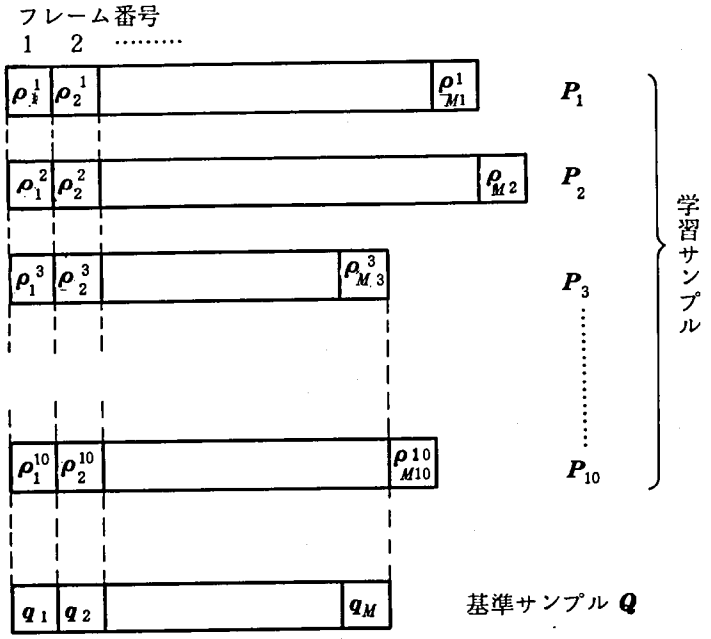


図 4.5　基準サンプルの作成

q_kの類似度を

$$l(\rho_i^j,\ q_k) \tag{4.21}$$

$(1 \leqq j \leqq 10,\ 1 \leqq i \leqq M_j,\ 1 \leqq k \leqq M)$

とする。次に，

$$L = \max\left\{\sum_{i=1}^{M_j} l(\rho_i^j,\ q_{f_j(i)})\right\} \tag{4.22}$$

を満足するf_jを求める。ただしf_jには次の条件が課せられているものとする。

$$\left.\begin{aligned} &f_j(1) = 1, \qquad f_j(M_j) = M \\ &f_j(i) = \begin{cases} f_j(i-1) \\ f_j(i-1)+1 \\ f_j(i-1)+2 \end{cases} \end{aligned}\right\} \tag{4.23}$$

この問題はDPを使って効率良く解くことができる。

(4)　上の操作をすべてのP_jについて行うと，関数の集合 $\{f_j\}$ $(j = 1, 2, \cdots\cdots, 10)$ が得

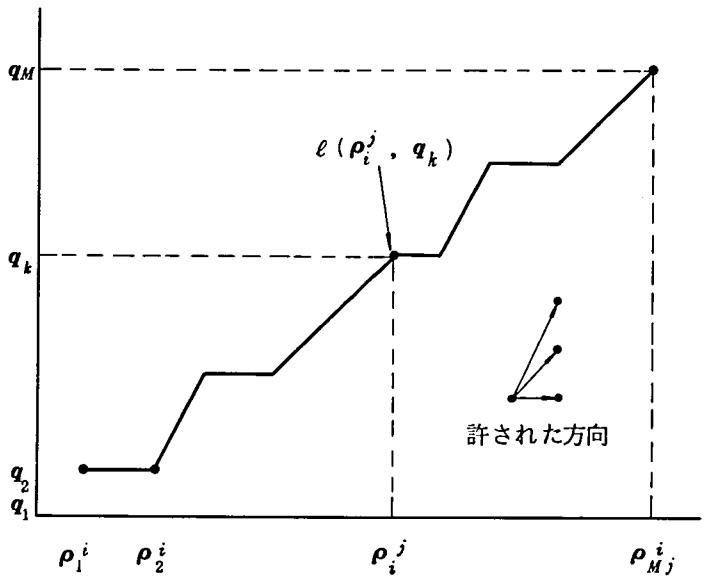


図 4.6　学習サンプルの時間軸の正規化

られる。各f_jは$\boldsymbol{P}_j$中の各特徴ベクトルを$\boldsymbol{Q}$の特徴ベクトルに対応づける写像関数と考えることができる。いま，q_kに対応づけられた特徴ベクトルの集合を$P(k)$とする。すなわち，$P(k)$は次式で表すことができる。

$$P(k)=(\rho_i{}^j \mid f_j(i)=k,\ 1\leqq i\leqq M_j,\ j=1,\ 2,\ \cdots\cdots,\ 10)$$
$$(k=1,\ 2,\ \cdots\cdots,\ M) \tag{4.24}$$

(5)　$P(k)$に含まれる特徴ベクトルを平均して$\boldsymbol{r}_k$を得る。すなわち，

$$\boldsymbol{r}_k=\frac{\sum_{\rho_j{}^i\in P(k)}\rho_j{}^i}{\mid P(k)\mid} \tag{4.25}$$

（$\mid P(k)\mid$は集合$P(k)$の要素の数）

この結果得られるM個のベクトルの系列，

$$\boldsymbol{R}=(\boldsymbol{r}_1,\ \boldsymbol{r}_2,\ \cdots\cdots,\ \boldsymbol{r}_M) \tag{4.26}$$

を標準パターンとする。（図 4.7）

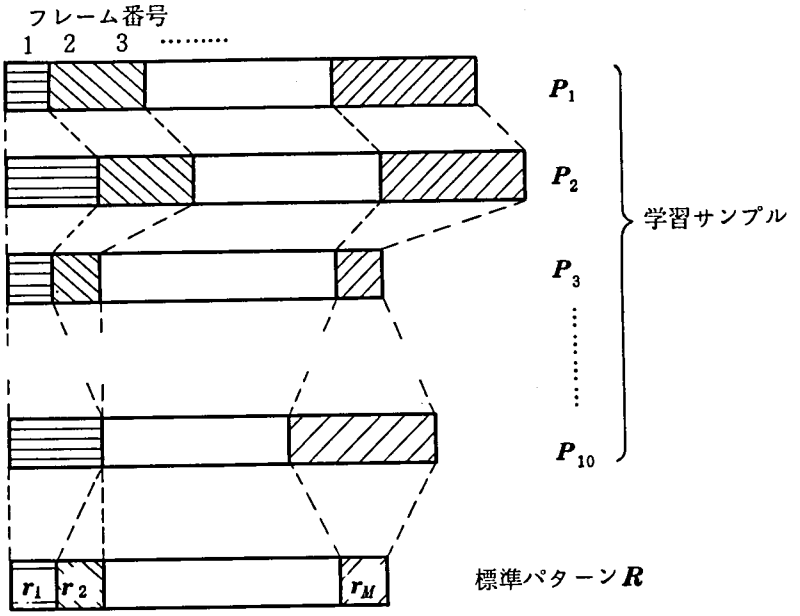


図 4.7　標準パターンの作成

4.4.2.2　認識方法

セグメンテーションされ認識部へ入ってくる情報は相関関数の時系列

$$\boldsymbol{P} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \cdots\cdots, \boldsymbol{\rho}_N) \tag{4.27}$$

パワーの系列

$$P = (v_{01}, v_{02}, \cdots\cdots, v_{0N}), \tag{4.28}$$

および，セグメント境界の組の集合

$$G = \{ G_1, G_2, \cdots\cdots, G_w \} \tag{4.29}$$

である。認識部の構成を図 4.8 に示す。図にそって認識の手順を説明する。

(1)　式（4.29）で与えられた集合から任意のG_jを選ぶ。このとき，

$$G_j = (j_1, j_2, \cdots\cdots, j_{e-1}) \tag{4.30}$$

であるとする。ただし，各j_iはセグメント境界のフレーム番号である。

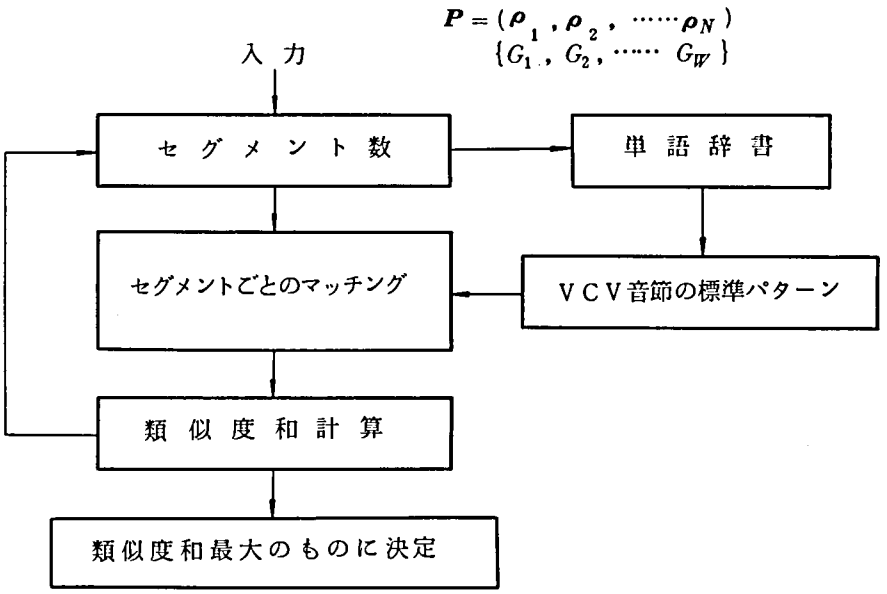


図 4.8　認識部の構成

(2)　単語辞書中には，認識対象となる単語がVCV音節の系列として蓄えられている。(たとえば，／ahiru／という単語は，／ahi／，／iru／という2つのVCV音節より構成されているという情報が入っている。) e 個のVCV音節で構成されている単語を探し，存在すればその単語を構成しているVCV音節の標準パターンを用意する。これを

$$\boldsymbol{R}_1, \quad \boldsymbol{R}_2, \cdots\cdots, \boldsymbol{R}_e \tag{4.31}$$

とする。

(3)　入力をセグメントの境界で切り，e 個の小系列

$$\boldsymbol{P}_1=(\rho_1, \rho_2, \cdots\cdots, \rho_{j_1}), \quad \boldsymbol{P}_2=(\rho_{j_1+1}, \cdots\cdots, \rho_{j_2}) \cdots\cdots$$
$$\boldsymbol{P}_e=(\rho_{je_{-1}+1}, \cdots\cdots, \rho_N) \tag{4.32}$$

を得る。(図4.9)

(4)　次に，$\boldsymbol{P}_j$ と $\boldsymbol{R}_i \equiv (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots\cdots, \boldsymbol{r}_{m_i})$ のマッチングを行う。一般に，$\boldsymbol{P}_j$ のフレーム数 $j_i - j_{i-1}$ と $\boldsymbol{R}_i$ のフレーム数 m_i は異なるので，標準パターンを作成する際と同じように時間軸の伸縮を補正したマッチッグを行う。まず，ρ_j と $\boldsymbol{r}_k$ の類似度を

$$l(\rho_j, \boldsymbol{r}_k) \tag{4.33}$$

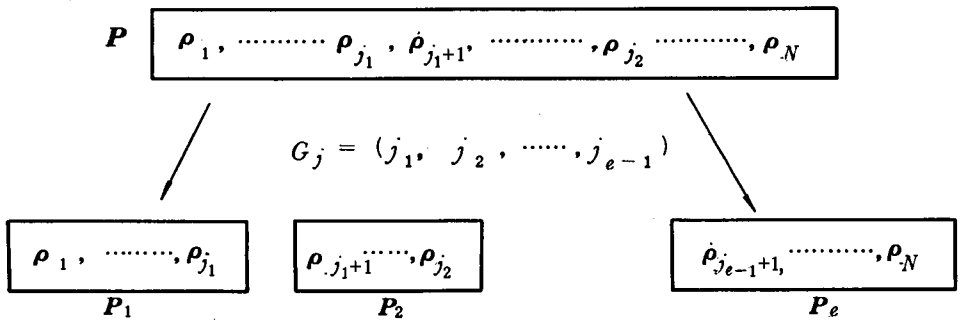


図 4.9　入力音声Pのセグメンテーション

とする。次に，

$$L_i = \max_f \left\{ \sum_{j_{i-1}+1}^{j_i} l\left(\rho_j,\ r_{f(j)}\right) \right\} \qquad (4.34)$$

を満足するfを求める。ただし，fには次の条件が課せられているものとする。(図 4.10)

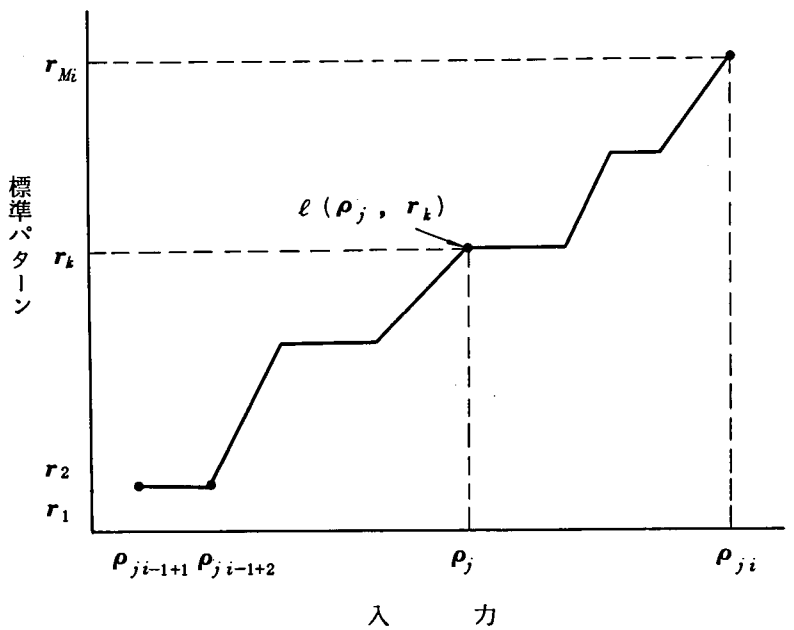


図 4.10　P_iとR_iのマッチング

$$\left.\begin{aligned} & f(j_{i-1}+1)=1, \quad f(j_i)=m_i \\ & f(j)=\begin{cases} f(j-1) \\ f(j-1)+1 \\ f(j-1)+2 \end{cases} \end{aligned}\right\} \qquad (4.35)$$

この問題を解いて得られるL_iが$\boldsymbol{P}_i$と$\boldsymbol{R}_i$の類似度になる。

(5) 各セグメントごとに標準パターンとのマッチングを行い，得られた類似度の和

$$\sum_{i=1}^{e} L_i \qquad (4.36)$$

を，入力と比較の対象となる単語の間の類似度とする。

(6) 同じ長さの他の単語に対しても上の操作を行い，得られた類似度を記憶しておく。同様に他のセグメント境界の組G_jに対しても同じ操作を行う。最後に，得られた類似度のうち最大の類似度を持つ単語が入力の属するカテゴリーであるとする。

4.4.2.3 パワー情報の利用

自己相関係数で表現された音声はスペクトルパターンに関する情報しか持っておらず，その他の重要な情報，たとえばパワー情報とか，ピッチ情報が含まれていない。したがって，入力音声の各セグメントの認識を行う際，これらの情報を有効に使えば，認識率の向上を計ることができると考えられる。第3章では，子音の前に休止区間が存在するかどうかで，あらかじめ子音の分類を行い，VCV音節の認識率が7%近く改善できることを示した。そこで，ここでも同じ方法をとる。すなわち，入力音声の各セグメントの認識を行う際，あらかじめ図4.11に示した判定論理により，子音の分類を行うことにする。図4.11に示した判定論理が第3章で用いた判定論理と異なっているのは，子音の前における休止区間の有無や長短が発声者によってかなり大きく変動することを考慮したためである。

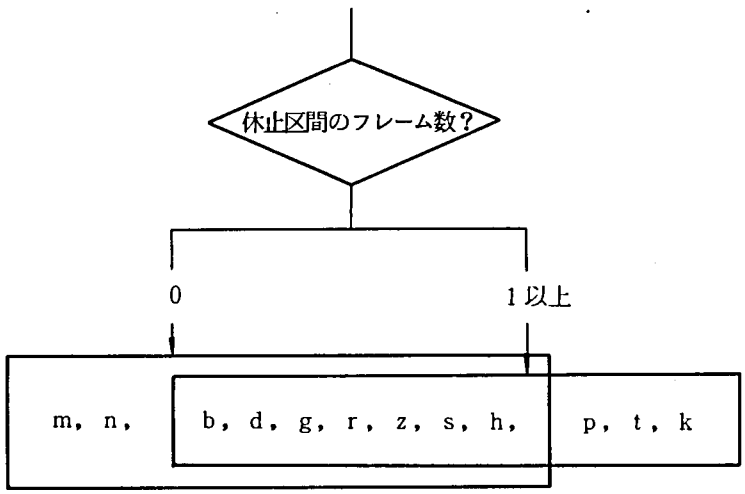


図 4.11　子音を分類する判定論理

4.5　認識実験[23][118]

4.5.1　セグメンテーション

4.5.1.1　セグメンテーションの閾値

まず，セグメンテーションの際の閾値を定める必要がある。いくつかの予備実験を行うことにより，閾値は次のように定めた。

$$\left.\begin{array}{ll} S_1^m = 3, & S_1^M = 30 \\ S_2^m = 10, & S_2^M = 40 \\ S_3^m = 20, & S_3^M = 50 \\ S_4^m = 30, & S_4^M = 60 \\ & \\ S_1 = 10, & S_2 = 20 \\ & \\ P_T = 10.0, & Q_T = 40.0 \end{array}\right\} \qquad (4.37)$$

これらの閾値の関係を図 4.12に示す。セグメンテーションの段階での脱落誤りをおさえるため，母音数の判定，休止，無声化の判別をする閾値は重なる部分をもうけてある。

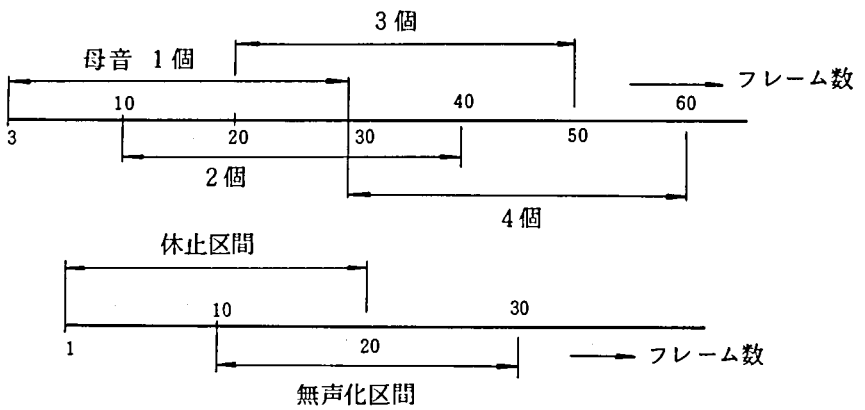


図 4.12　セグメンテーションの際の閾値

4.5.1.2　セグメンテーションの実験

先に述べた8名の発声者による3,580サンプルを4.4.1で述べた方法によりセグメンテーションした。STEP 1の操作で得られた区間のフレーム数と中に含まれる母音数の関係をグラフにして図4.13に示す。また，休止区間と無声化区間のフレーム数の分布の様子を図4.14に示す。これらのサンプルに4.5.1.1で定めた閾値を適用した場合いくつかの誤りが生じるが，これは認識の段階で回復不可能な誤りである。くわしいことは検討の項で述べる。

4.5.2　単語認識

4.5.2.1　個人別，単語別の認識率

4.5.1と同じサンプルをセグメンテーションの後，認識部へ入力することにより単語認識を行った。認識結果を，単語の種類別，個人別にまとめて，表4.2に示す。平均の認識率は96.5％である。

4.5.2.2　鼻音化の対策

発声者によっては，／g／が単語中で鼻音化して／$\tilde{g}$／の音になる傾向がある。しかしながらVCV音節の標準パターンとしては／$\tilde{g}$／を含んだものは用意していないため，鼻音化の傾向の

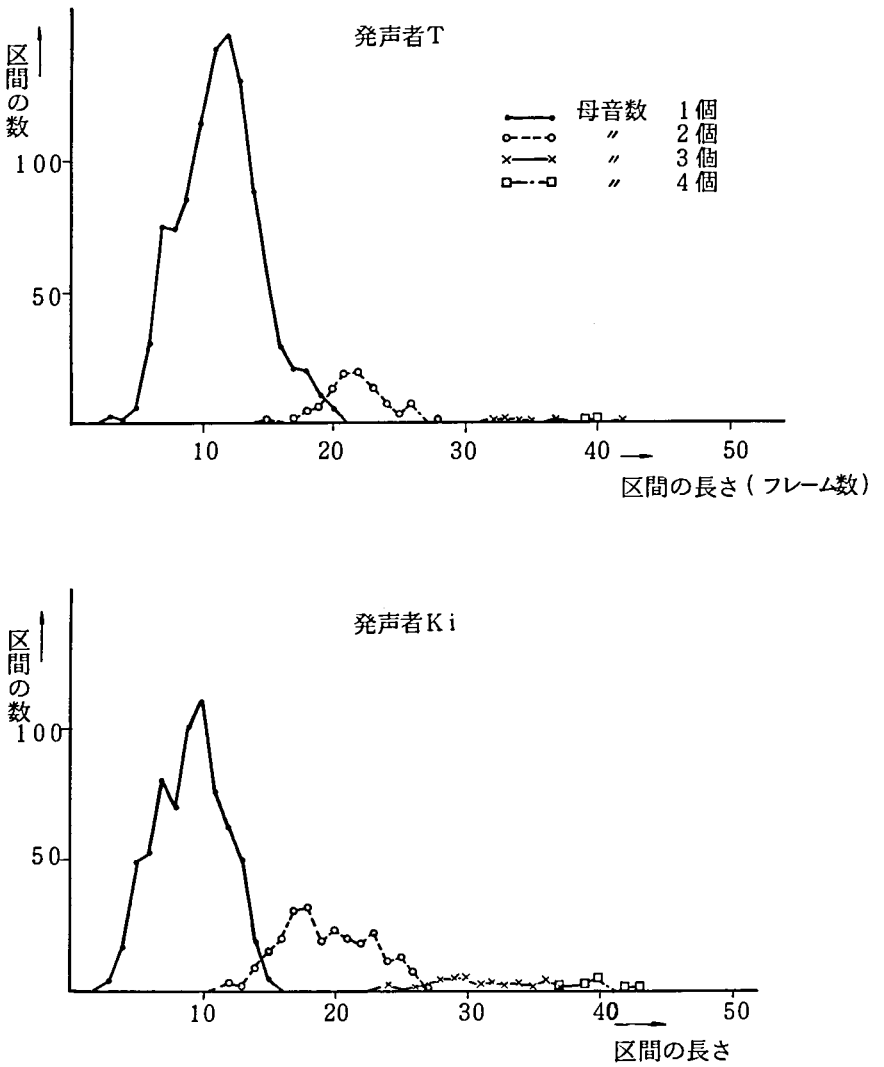


図 4.13　区間の長さと中に含まれる母音数の関係 (1)

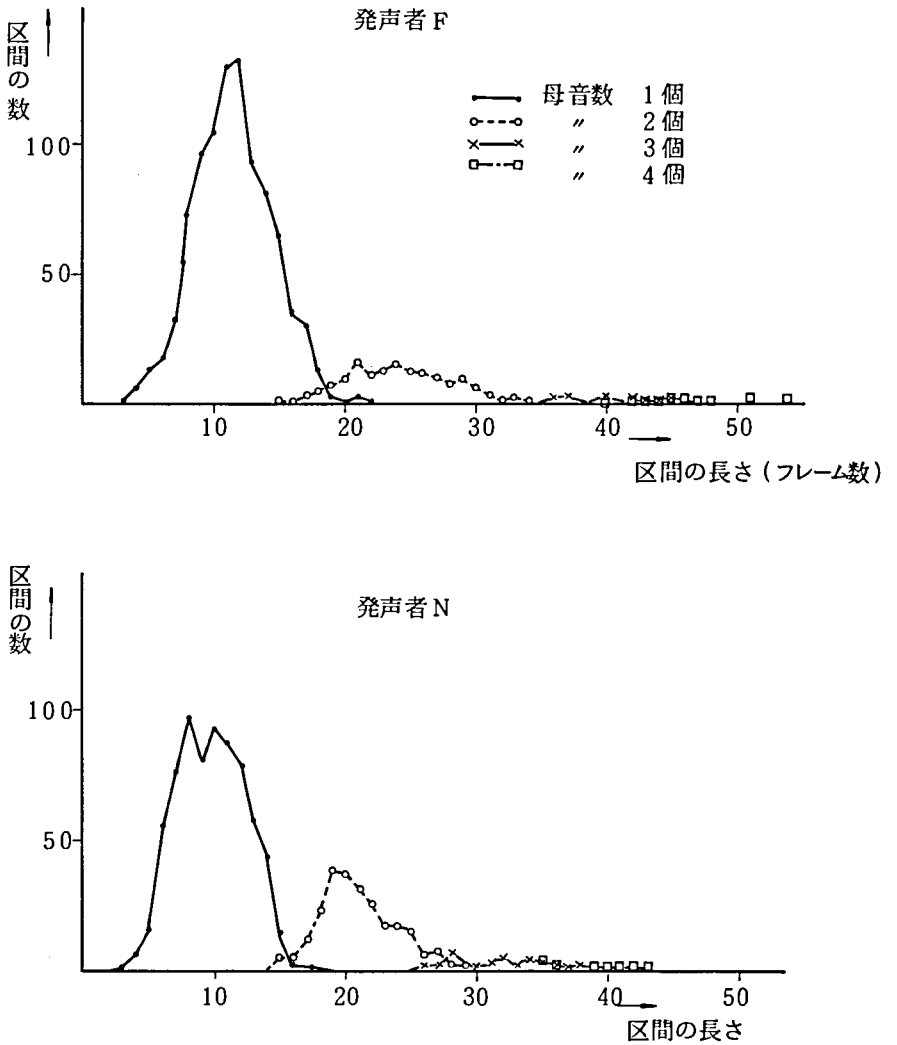


図 4.13　区間の長さと中に含まれる母音数の関係 (2)

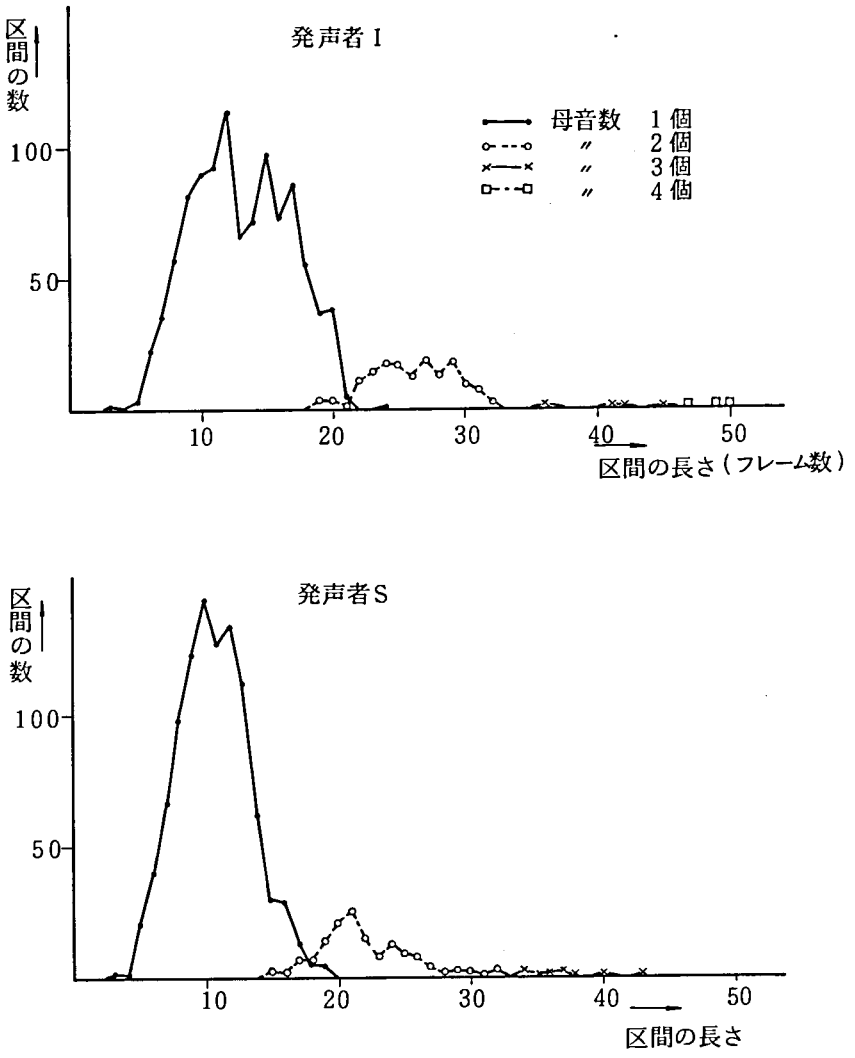


図 4.13　区間の長さと中に含まれる母音数の関係 (3)

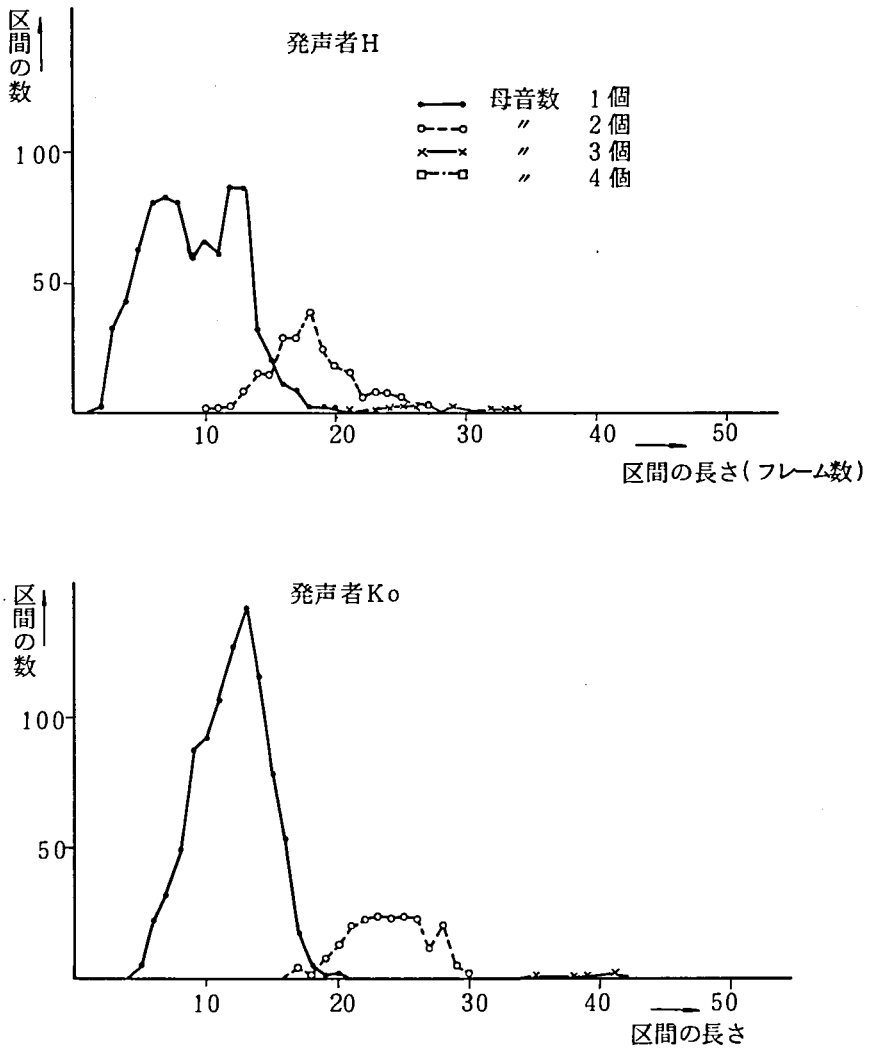


図 4.13　区間の長さと中に含まれる母音数の関係(4)

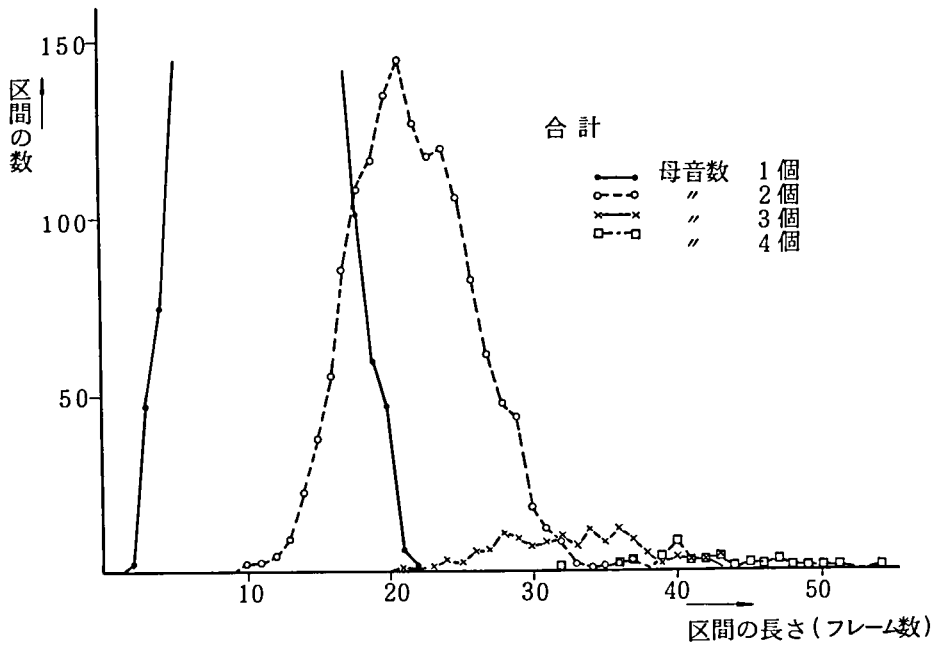


図 4.13　区間の長さと中に含まれる母音数の関係 (5)

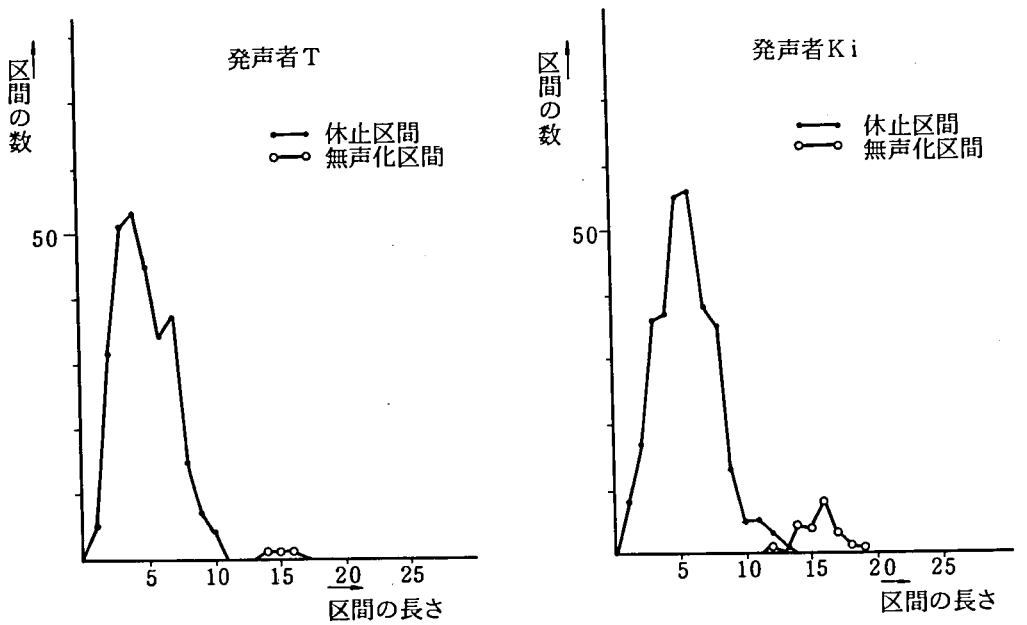


図 4.14　休止区間と無声化区間の区分数の分布 (1)

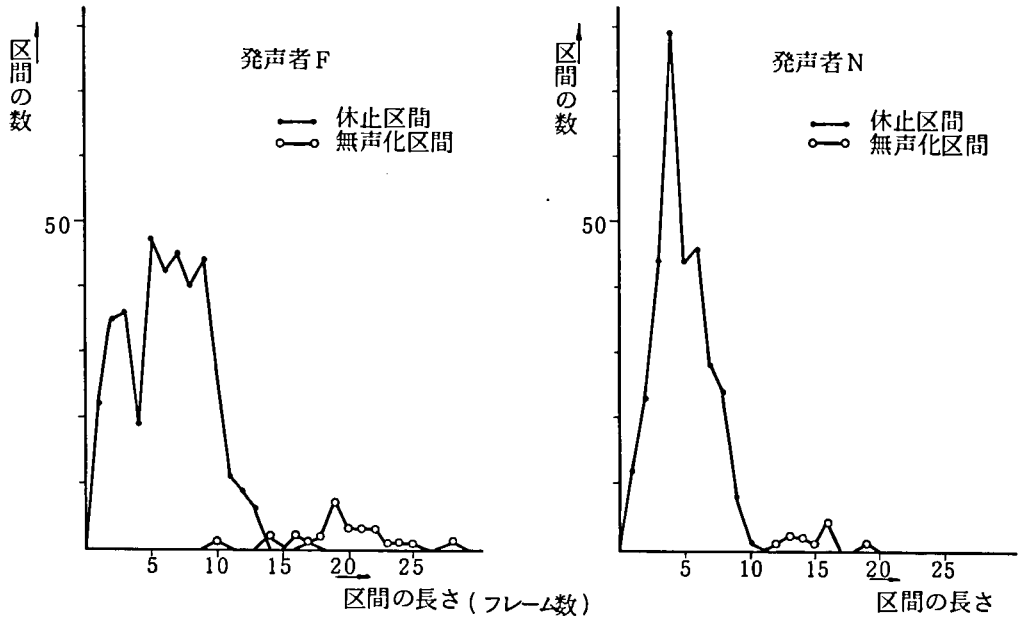


図 4.14　休止区間と無声化区間の区分数の分布 (2)

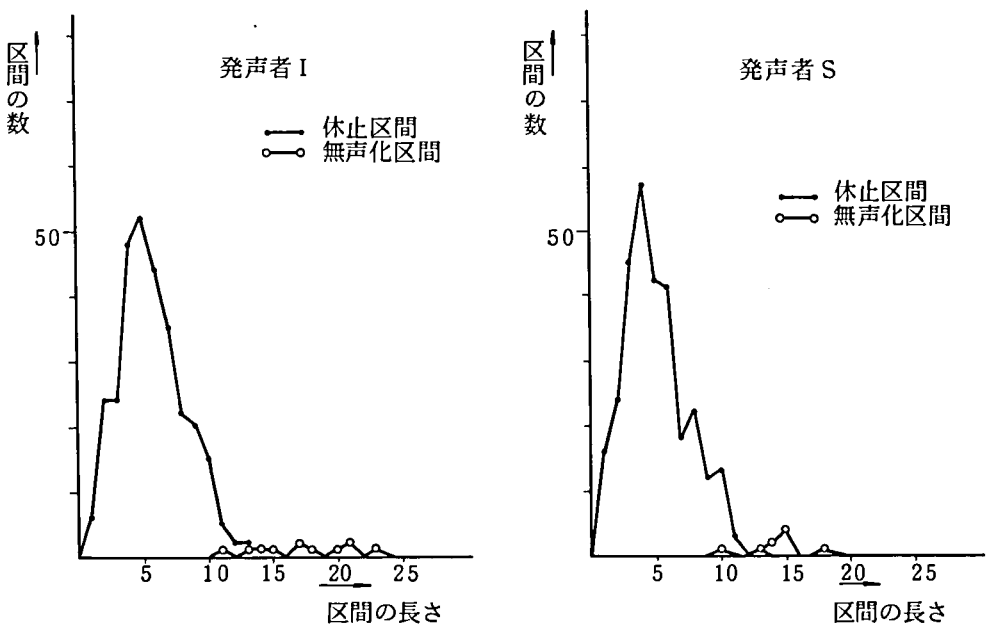


図 4.14　休止区間と無声化区間の区分数の分布 (3)

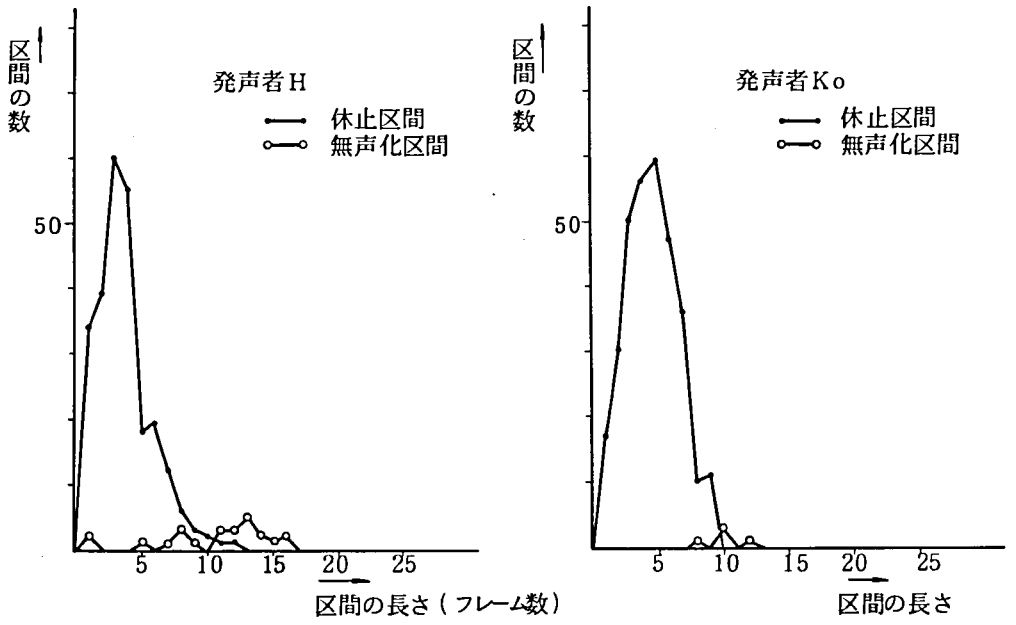


図 4.14　休止区間と無声化区間の区分数の分布 (4)

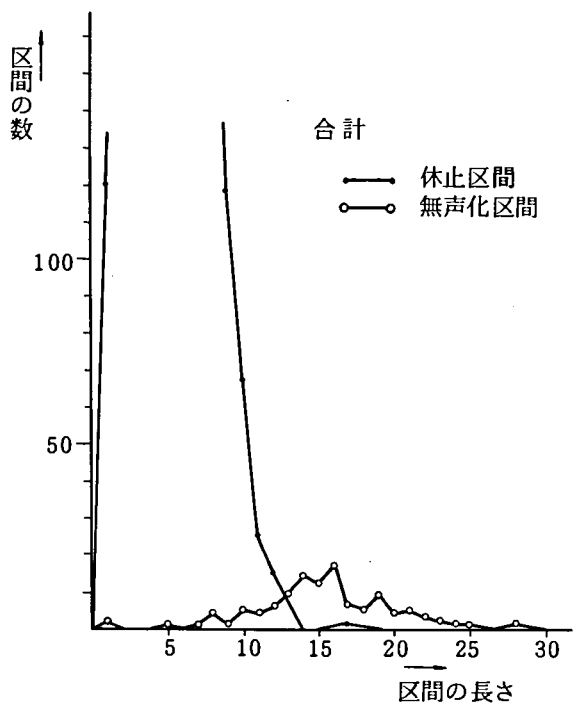


図 4.14　休止区間と無声化区間の区分数の分布 (5)

表4.2　認　識　結　果

単語の種類 / 発声者	V C V 型	V C V C V 型	VCVCVCV型	計
T	98.0 % 147/150	100 % 150/150	100 % 150/150	99.3 % 447/450
Ki	94.7 % 142/150	95.3 % 143/150	100 % 144/144	96.6 % 429/444
F	90.0 % 135/150	95.3 % 142/149	99.3 % 149/150	94.9 % 426/449
N	95.3 % 142/149	96.6 % 144/149	98.7 % 147/149	96.9 % 433/447
I	92.0 % 138/150	98.7 % 148/150	100 % 150/150	96.9 % 436/450
S	96.7 % 145/150 (100 % 150/150)	94.7 % 142/150 (99.3 % 149/150)	98.7 % 148/150 (98.7 % 148/150)	96.7 % 435/450 (99.3 % 447/450)
H	96.6 % 143/148 (99.3 % 147/148)	95.3 % 141/148 (95.3 % 141/148)	87.5 % 126/144 (89.6 % 129/144)	93.2 % 410/440 (94.8 % 417/440)
Ko	94.7 % 142/150	98.7 % 148/150	98.0 % 147/150	97.1 % 437/450
計	94.7 % 1134/1197 (95.5 % 1143/1197)	96.8 % 1158/1196 (97.4 % 1165/1196)	97.8 % 1161/1187 (98.1 % 1164/1187)	96.5 % 3453/3580 (97.0 % 3472/3580)

（かっこの中は鼻音化の対策を施した場合の認識率）

著しい発声者（発声者S，H）では，このための誤認識が多い。そこで，発声者S，Hに対し次のような簡単な鼻音化の対策を施して認識率の向上をはかった。すなわち，／g̃／の音が／n／に近いことから，単語辞書中には／g／を含んだ単語に対しては／g／を／n／でおきかえたものも用意しておく。たとえば，／igo／，／udegumi／等の単語に対しては，／ino／，／udenumi／等も単語辞書中に用意しておく。そして，認識の場合には，どちらに判定されても正しく認識されたものとする。このような対策を施した場合の認識率を表4.2のカッコの中に示す。このとき，全体の認識率は97.0％に向上する。

4.6 検　討

4.6.1 認識率の個人差，単語の種類による認識率

表4.2に示した，個人別，単語の種類別の認識率をわかりやすくグラフにして図4.15および図4.16に示す。ただし，図4.15は鼻音化の対策を施す前後の認識率の変化を発声者S，Hおよび全体の平均について示してある。認識率は発声者によってかなりの差があるが，認識率の良し悪しは必ずしも人間が聞いた場合の声の良し悪しとは対応していない。これは，セグメンテーションの際の閾値が，発声者によって適合性に差があることや，標準パターン作成用のサンプルと認識用のサンプルでは発声条件が異なる場合がある（たとえば風邪をひいていた等）ことなどでいくらかは説明できよう。しかしながら，音声の個人差の問題はまだ解明されていない点が多く，このような問題点の説明には今後の研究に待つところが多い。

次に，単語の種類別でみると，単語が長くなると認識率が良くなる傾向が見られる。これは，単語が長くなれば考えうる単語の種類が多くなるために，認識対象が全単語の中で占める割合が小さくなるからである。言い換えれば辞書の効果ということができよう。発声者によっては逆に単語が長くなると認識率の下るものもいるが，これは後で述べるようなセグメンテーションの誤りが多く生じているためである。

4.6.2 誤りの分析

単語認識における誤りの原因は，セグメンテーションの誤りとVCV音節の認識誤りに分類される。発声者別に誤りの数を分類したものを表4.3に示す。以下それぞれの誤りについて説明する。

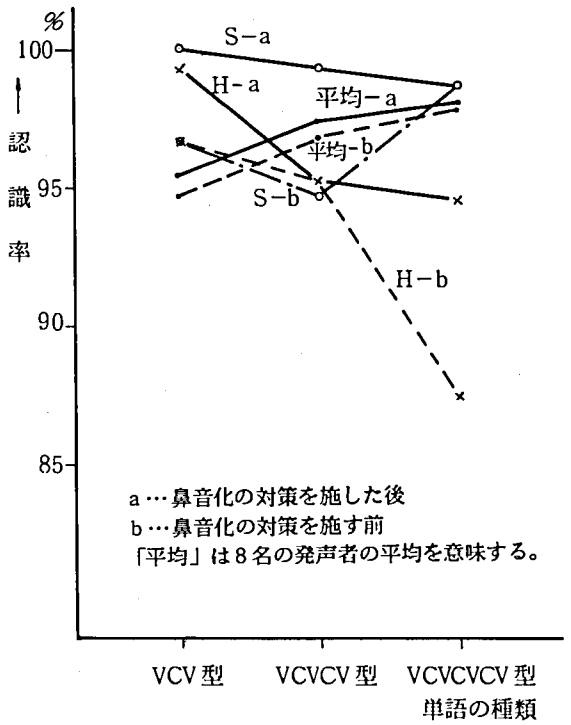


図 4.15　鼻音化の対策を施す前後の認識率の比較
（発声者S，Hおよび平均について）

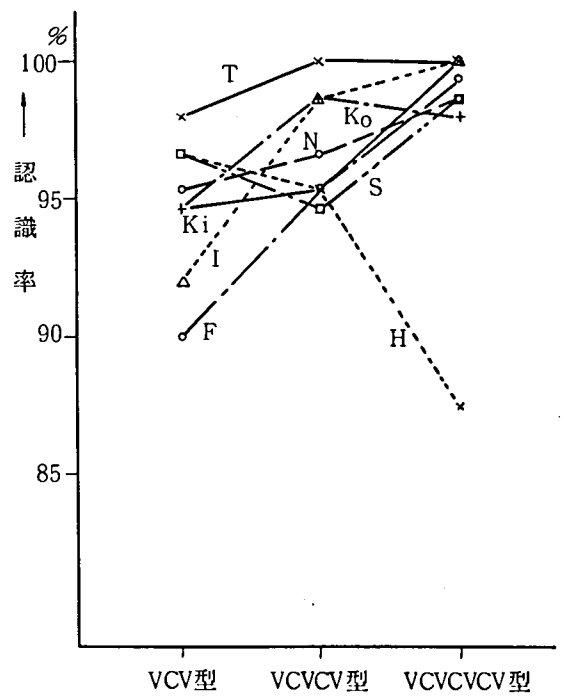


図 4.16　個人別，単語の種類別の認識率

表4.3 誤りの分類

誤りの種類＼発声者			T	K_i	F	N	I	S	H	K_o	計
セグメンテーションの誤り	母音数決定の誤り								3		3
	無声化区間→休止区間								9	1	10
	子音の検出				3			2		3	8
VCV音節の認識誤り	同じセグメント数の単語間のconfusion	VCV →VCV	3	8	14	7	11	6 (0)	5 (1)	7	61 (51)
		VCVCV →VCVCV			3	2	2			1	8
		VCVCVCV →VCVCVCV				2			1		3
	異なったセグメント数の単語間のconfusion			7	3	3	1	7 (1)	12 (9)	1	34 (25)
計			3	15	23	14	14	15 (3)	30 (23)	13	127 (108)

（かっこの中は鼻音化の対策を施した場合）

4.6.2.1 セグメンテーションの誤り

ここでいうセグメンテーションの誤りとは，セグメンテーションの段階で生じ，認識の段階では回復不可能な誤りのことである。このような誤りは表4.3から解るように計21個生じた。

(1) 母音数決定の誤り……区間長と母音数の関係を示す閾値が個人差や発声速度を完全にはカバーできなかったために生じた誤りである。この誤りを減少させるには，区間中に含まれる母音数を規定する閾値の幅を広げればよいが，そのようにするとセグメンテーションの候補数が多くなり，認識の段階での誤りが生じやすくなる。8人の発声者で3個の誤りであるから，ここで用いた閾値は，現在のままでも十分適用範囲が広いと考えられる。

(2) 無声化区間が休止区間と判定された誤り……無声化の判定を行う閾値S_2の設定が不十分だったために生じた誤りである。この誤りは発声者Hに集中して生じており，無声化区間のフレーム数が発声者によってかなり変動しやすいものであることがわかる。誤りの生じたサンプルについてソナグラムを見て調べてみると，無声化といっても次に示すような種々の現象が生じていることがわかる。

a. 母音の子音化……母音が発声されず，代りに直前の子音がかなりのパワーを伴ってある程度の継続時間発声されるもの。

b. 母音の無音化……母音が発声されず，長い無音区間が続くもの。

c. 母音の脱落……b.で無音区間が非常に短くなったもの。無声化した母音の前後の子音が連続して発声されたように聞こえる。

このように，母音の無声化はかなり複雑な問題であり，単に無音区間の長短で判定するのは危険であると思われ，より詳細な検討が必要である。

逆に，休止区間を無声化区間と判定した誤りは生じていない。したがって，閾値S_1の値は妥当といえる。

(3) 子音の検出による誤り……図4.17に示したように，／s／，／h／のような子音はかなり大きなパワーを持つことがあり，そのため母音と判定される場合が生じる。このような原因による誤りは表4.3に示したように8個生じている。これは，単にパワーの大小で母音，子音の判定を行っているために生じた誤りであるから，スペクトル情報等より詳細な情報を用いることが必要である。

4.6.2.2 VCV音節の認識誤り

4.6.2.1で述べた誤り以外の誤りは，セグメントの認識が十分正確に行えなかったことが主

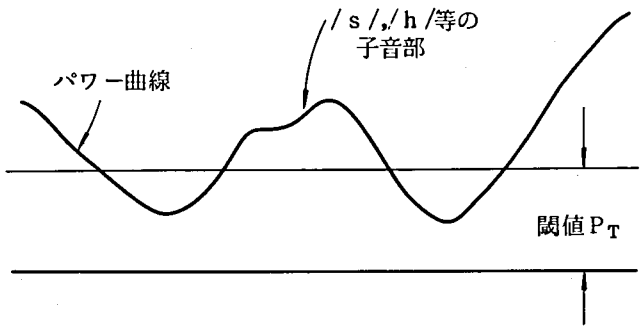


図 4.17　子音の検出

な原因で生じたものである。これらの誤りについていくつかの検討を加える。

(1)　VCV型→VCV型なる誤りは全体の 48.0％（鼻音化対策を施した後では 47.2％）を占めている。表 4.4 に confusion matrix を示す。／iro／→／igo／，／eri／→／ebi／のように，本来かなり識別が困難な単語間の誤りが多いことがわかる。VCV型単語はそのままVCV音節であるから，今後認識系の性能を上げるには，このようなよく似たVCV音節を正しく識別する手段を考える必要がある。

(2)　単語の各セグメントがどの程度正しく認識されているかを調べるために，代表的な発声者 1 名（発声者T）について，セグメンテーションされた各VCV音節の認識を行った。結果を表 4.5 に示す。この結果から解るように，単語が長くなると単語中のVCV音節の認識率は下る傾向がある。これは，連続音声中のVCV音節が，調音結合の影響を受け，単独に発声されたものに比べ変形するためと考えられる。100 語程度の単語数を対象にした場合は，辞書の効果がこのような欠点を補ってくれるが，より多い語を対象とした連続音声の認識を目ざす場合には，このような現象を解明し，対策をこうじる必要がある。

(3)　VCV音節の認識誤りに分類した誤りの中にも，セグメンテーションの方法を改良することにより救うことのできるものが少なくない。それは次のような理由による。

- セグメンテーションの候補は後で述べるように，単語 1 個当り 2.7 個である。このようにセグメンテーションの段階で候補をしぼることができなかったことが誤りの生じた原因の 1 つである。特に異なった音節数の単語間の confusion はこの原因によるものが多い。
- セグメンテーションの手順STEP 3 ではセグメントの境界を機械的に定めるという方法をとっているため，母音の定常部からずれた部分をセグメントの境界としている恐れがあり，そのための誤りが生じている可能性がある。

表 4.4　VCV型単語の confusion matrix

（鼻音化の対策を施した場合）

入力＼出力	歌	おば	駅	医務	後	鬼	色	尾根	あぜ	いす	一	奥	裏	汗	居間	えさ	囲碁	仇	牛	音	えり	海老	F	梅	右派	味	江戸	池	オノ	異議	その他	計
歌	40																															40
おば		40																														40
駅			40																													40
医務				40																												40
後					40																											40
鬼						39					1																					40
色							30										10															40
尾根								40																								40
あぜ									39					1																		40
いす										40																						40
一			1								35																			4		40
奥												39																				39
裏													39												1							40
汗														40																		40
居間															40																	40
えさ																40																40
囲碁							7										33															40
仇																		39													1	40
牛																			39												1	40
音																				40												40
えり																					18	22										40
海老																					2	37								1		40
F																							38									38
梅																								40								40
右派																									40							40
味																										39					1	40
江戸																											40					40
池																												40				40
オノ		1																											39			40
異議																														40		40

表 4.5　単語中のVCV音節の認識結果（発声者 T）

（母音部分はわかっているものとして認識実験をおこなったので子音の認識結果のみを示す）

VCV型単語中のVCV音節

入＼出	m	n	b	d	g	r	z	p	t	k	s	h
m	15											
n		13			1	1						
b			9	1								
d				9			1					
g					10							
r			1	4		10						
z							10					
p												
t									20			
k										15		
s											15	5
h												10

VCVCV型単語中のVCV音節

入＼出	m	n	b	d	g	r	z	p	t	k	s	h
m	18	6	1									
n	1	27			1	1						
b			26	2			2					
d			1	13	1							
g			1	5	19							
r	1	3	3	3	5	28	2					
z			1				14					
p								5				
t								3	21	1		
k									2	38		
s									6		14	10
h									1			14

VCVCVCV型単語中のVCV音節

入＼出	m	n	b	d	g	r	z	p	t	k	s	h
m	33	5	11	1								
n	3	26		4	1	1						
b			24	5	1							
d			1	11	2		6					
g		1	2	2	42	1	2					
r	2	2	5	14	6	20	1					
z				9			21					
p								10				
t							2	5	43			
k					4			6	1	54		
s				1							18	11
h	1	1				2	7			1		18

VCV型単語中の VCV音節	VCVCV型単語中の VCV音節	VCVCVCV型単語中の VCV音節	計
136/150	237/300	320/450	693/900
90.7 %	79.0 %	71.1 %	77.0 %

4.6.3 鼻音化の対策の効果

発声者S，Hの誤りのうち，/g/の鼻音化によって生じたと考えられる誤りのリストを表4.6に示す。これらの20個の誤りのうち，1個を除いて鼻音化の対策を施すことにより正しく認識できた。ただし，認識対象がより増加する場合には /g̃/の標準パターンを用意しておく必要があろう。また，この他にも変化しやすい音韻があるならば，その規則性を見出し，音韻レベルの情報として蓄えておくことが必要になる。

表4.6　鼻音化による誤りのリスト

発声者	入　力	認識結果	
S	*igo*	*ima*	(4個)
	igo	*iro*	
	igi	*imu*	
	otsuge	*one*	(2個)
	enogu	*ono*	(2個)
	enogu	*edo*	
	udegi	*udegumi*	
H	*igo*	*iro*	(4個，うち1個は鼻音化の対策によって救えず)
	igo	*imu*	
	agohige	*umibe*	
	anaguma	*abura*	

4.6.4 セグメンテーションの候補数

表4.7に単語1個当りのセグメンテーションの候補数を発声者別，単語の種類別にして示した。各発声者共，単語1個当り2.4～3.1個の候補が生じていることがわかる。このように，多くの候補があげられることは，先に述べたように単語認識の際の誤りの原因となる可能性があり，また計算時間の上でも実際に必要な時間の2.4～3.1倍の時間を要していることになり好ましくない。したがって，セグメンテーションの際，パワー以外の情報も使ってできるだけ候

補をしぼる必要がある。

表4.7　単語1個当りのセグメンテーション結果の候補数

単語の種類 / 発声者	ＶＣＶ型	ＶＣＶＣＶ型	VCVCVCV型	平　　均
T	1.59	3.34	2.81	2.58
K_j	2.18	2.66	2.49	2.44
F	3.29	3.24	2.31	2.95
N	1.95	2.79	2.64	2.46
I	2.61	3.48	2.69	2.92
S	1.73	3.28	2.51	2.51
H	1.74	2.94	3.18	2.62
Ko	2.83	3.88	2.65	3.12
平　　均	2.24	3.20	2.66	2.70

4.6.5　今後の問題点

最後に，以上述べてきた問題をまとめておく。

4.6.5.1　セグメンテーションの問題点

ここで採用したセグメンテーションの方法は，単語認識を行う限りでは，十分応用範囲の広いものと考えられる。より誤りを少なくするには次のような点を改良する必要がある。

(1)　セグメンテーションの候補をできるだけしぼるようにする。

(2)　機械的にセグメントの境界を定めるのでなく，他の情報を使って母音定常部の中央をセグメントの境界に決定するようにする。

(3)　無声化の判定の方法をより精密化する。

(4)　母音部，子音部の判別にもパワー以外の情報を用いるようにする。

4.6.5.2　ＶＣＶ音節認識の問題点

誤りの多くはVCV音節の認識誤りに起因している。しかしながら，VCV音節単位の認識率は子音の認識率としてはかなり高いものであり，VCV音節を音声単位にとることが有効であるこ

とを示している。今後の問題点として次のものが考えられる。

(1) 連続音声中でVCV音節が調音結合の影響を受けて変形する場合に認識率が低下しないようにする必要がある。

(2) スペクトルの形状に関する情報以外の情報を積極的に使って認識率の向上をはかる必要がある。

4.7 あとがき

以上述べてきたように，本章では，VCV音節を単位として単語音声を認識するシステムを構成し，90種類の日本語単語を認識対象として認識実験を行った。その結果，8名の発声者による3,580サンプルを用いて97.0％の認識率が得られた。これは，単語より小さい単位を認識の単位とする認識方式としては十分高い認識率である。また，さらに認識率を上げるには，セグメンテーションの精密化，連続音声中のVCV音節の認識率向上が必要であるという問題点を明らかにした。これらの問題点の検討は次章以降で行う。

第5章　ＶＣＶ音節を単位とした連続単語音声の認識

5.1　はしがき

第4章では，VCV音節を単位とした単語音声の問題を取り扱った。本章では対象を連続単語音声に拡大し，VCV音節を単位として，連続単語音声を認識する手法，および，認識実験について述べる。連続単語音声を認識対象とすることの意義は次の点にある。

(1)　単語を1語1語，区切って発声することは，発声者にかなり負担をかけることになり，情報伝達速度も遅くなるため，実用的立場からは，単語を連続して発声することを許した連続単語音声の認識が可能であることが望ましい。

(2)　連続単語音声の認識においては，単語音声の場合に比較すると，単語単位のセグメンテーションという新たな問題が生じる。この問題は，会話音声認識にも共通した問題であるため，連続単語音声は，単語音声から会話音声へ研究を進める際の中間的な目標として適当である。

以上の理由に基づき，本章では，日本語から選んだ90種類の単語を認識対象とし，これらの単語を連続して発声した音声を認識することを試みる。

5.2　連続単語音声の認識方針[70]

連続単語音声を認識するには，入力音声の単語境界の検出，すなわち，単語単位のセグメンテーションを行う必要が生じる。音響レベルでは，単語境界は明確な物理量としては現われないため，セグメンテーションを単独で行うことはできない。従って，認識とセグメンテーションを並行して行いながら，処理を進めていく必要がある。このような処理を簡単に行う方法として，迫江，千葉が提案した方法[34]がある。これについて説明する。入力音声をＡとする。標準パターンをＡの先頭から時間軸の非線形な伸縮の補正を行いながら重ね合わせていく。この操作は，図5.1に示したように，標準パターン（B_0とする）とＡの類似度マトリクス上で，原点から第 l_0 行に至る類似度和が最大になるパスを求めることに対応する。最適のマッチングが得られた，すなわち，最大の類似度が得られた標準パターンをB_0とする。このとき，入力音声の第1語の認識は終了したものとし，認識結果をB_0とする。また，入力音声は l_0 でセグメン

テーションされる。ℓ_0以前の部分を取り去ったパターンを新たにAとし，同様の処理を繰り返すことにより，順次，認識とセグメンテーションを行っていき，認識結果B_0，B_1，……を得る。

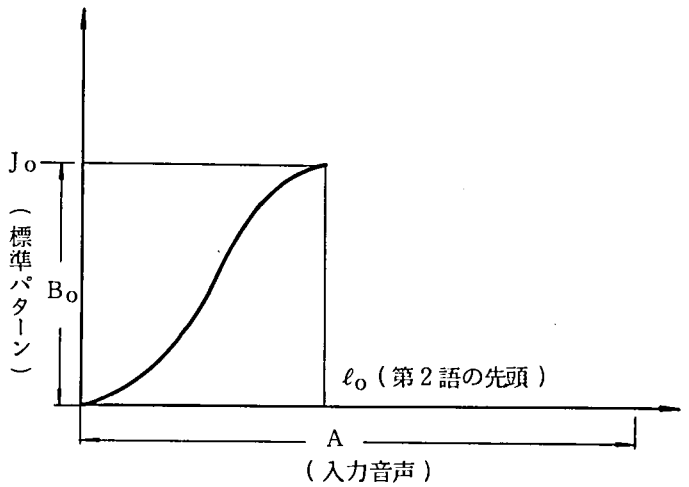


図 5.1　文献（34）の連続単語認識法

このような手法は，比較的少ない処理量ですむ利点はあるが，1語ずつ認識していくため，一度誤りが生じると次々に波及していき，回復が困難になる欠点を持っている。特に，会話音声と異なって，連続単語音声においては，構文情報，意味情報の補助がないため，1語の認識誤りはそれ以後の単語の認識にとって致命的である場合も多い。

このような欠点をなくし，高い認識率を得るためには，入力音声全体での最適な単語系列を決定するのが良いと考える。具体的には，すべての単語系列を作成し，それと入力音声との類似度を持つ単語音声を認識結果とする方法を採用する。このような方法では，当然，処理量は膨大になるが，ここでは，次のような処理方法をとることにより，処理量を少なくおさえることができた。

(1)　単語より小さいVCV音節を認識の単位にとったため，セグメンテーション，セグメンテーションされたVCV音節の認識，の段階で情報量が圧縮できる。

(2)　単語系列と入力音声の類似度を求める際，DPを用いる。DPは，前の処理結果を使いながら，一段づつ計算を進めてゆく手法である。例えば，n個の単語の系列と入力との類似度は，n－1個の単語の系列と入力との類似度，および，1個の単語と入力との類似度の和として求められる。順次nを増やして計算することにより，任意の長さの単語系列と入力との類似度が容易に求められる。従って，DPを用いれば，単語系列ごとに類似度を求める必要はなく，単語単位で類似度を求めればよいため，処理量が大幅に削減できる。

以上の考察に基づいた連続単語音声認識系の構成を図5.2に示す。認識系は音響処理部と連

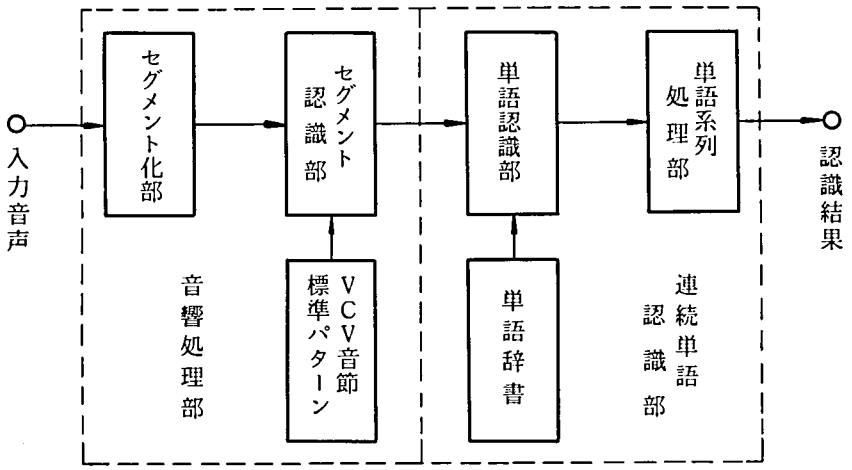


図5.2　連続単語認識系の構成

続単語認識部から構成されている。さらに音響処理部は，前処理部，セグメント化部，セグメント認識部から構成されており，連続単語認識部は，単語認識部，単語系列処理部から構成されている。入力音声は，前処理された後，VCV単位にセグメント化され，次に各セグメントの認識が行われて離散的な記号系列に変換される。連続単語認識部では，単語認識機能を持った単語認識部と，その出力結果からDPを用いて最適の単語系列を決定する単語系列処理部によって連続単語の認識が行われる。次節以後で各部についてくわしく説明する。

5.3　音響処理部[24][70]

音響処理部は前処理部，セグメント化部，セグメント認識部より構成されている。入力音声を離散的な記号の系列に変換する際には情報の脱落がおこりやすいので，それを防ぐため，セグメント化やセグメントの認識の際のあいまいさをある程度残しておいて結果を次段へ送るという構成をとる。

5.3.1　前処理部

2.4で述べたのと同じ方法で特徴抽出を行う。入力音声は3.2 kHz低域通過フィルタを通した

後，標本化周波数 8 kHz の AD 変換器で 11 ビットのディジタル音声に変換される。ディジタル音声は 15 msec ごとに区分され，音声パワーが求められる。あらかじめ閾値を定めておき音声パワーが閾値以上になったフレームを音声区間の始端とする。閾値以下の音声パワーが20フレーム（300 msec）以上続いたときは最初に閾値以下になったフレームを音声区間の終端とする。

以上の手続きにより音声区間が決定される。音声区間においては 15 msec のフレームごとに音声の特徴量として音声波形の自己相関関数

$$\boldsymbol{v} = (v_0, v_1, \cdots\cdots, v_p) \qquad (5.1)$$

を求める。したがって入力音声は $\boldsymbol{v}$ の時系列

$$\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\cdots, \boldsymbol{v}_N) \qquad (5.2)$$

およびパワーの系列

$$P = (v_{10}, v_{20}, \cdots\cdots, v_{N_0}) \qquad (5.3)$$

として表現される。ただし N はフレーム数であり，$\boldsymbol{v}_i$，v_{i0} は第 i フレームの自己相関関数およびパワーである。セグメント化部へは（5.2），（5.3）の情報が送られる。

5.3.2 セグメント化部

前章で述べた単語認識では，簡単な方法でセグメント化を行うという方針にそってパワーと継続時間の情報のみを使って音声から VCV 音節を抽出する方法を採用した。このような抽出法では確実な VCV 音節の抽出が困難であり，誤りを防ぐためにはかなりのあいまいさ* をゆるす必要があった。このように，セグメント化の段階であいまいさがかなり残っていると処理量が増大し，かつ認識誤りも生じやすい。したがって，連続単語音声のようにより複雑な音声を扱うには，セグメント化の方法を精密化して，あいまいさの減少をはかり，かつ誤りが生じることを防ぐ必要がある。以上のような考え方に従い，ここではセグメント化のための情報をふやすと共にセグメント化の操作を，(1)音韻境界の抽出，(2)音韻区間の分類，(3) VCV 音節の抽出，の 3 つの操作に分割することにより処理の精密化をはかった。

---

* 本論文中では「あいまいさ」という言葉を「複数の候補を持つ」という意味で用いる。以後の章でもこの言葉はしばしば用いるが，同じ意味である。

(1) 音韻境界の抽出

次の4つの情報を音韻区間抽出のために用いる。

a. パワー系列(5.3)

b. 急激なスペクトルの変化を示す系列……母音ー子音間の遷移のように急激にスペクトルが変化する時点を検出するための情報として，d_{ij}*を第iフレームと第jフレームのスペクトルの差としたとき次の量を用いる。

$$(e^1_2,\ e^1_3,\ \cdots\cdots,\ e^1_{N-2}) \qquad (5.4)$$

ただし，$e^1_i = (d_{i-1,\,i+1} + d_{i,\,i+2})/2$

c. ゆるやかなスペクトルの変化……母音間の遷移のようにゆるやかなスペクトルの変化を抽出するために次の量を用いる。

$$(e^2_3,\ e^2_4,\ \cdots\cdots,\ e^2_{N-3}) \qquad (5.5)$$

ただし，$e^2_i = (d_{i-2,\,i+2} + d_{i-1,\,i+3})/2$

d. 母音系列……母音標準パターンを用いて各フレームの認識を行い，得られた母音標準パターンは最尤スペクトルパラメータの形でたくわえられている。母音 /x/ の標準パターンを，

$$(A_{x0},\ A_{x1},\ \cdots\cdots,\ A_{xp}) \qquad (5.6)$$

---

* 第iフレーム，第jフレームのスペクトルを$P_i(w)$，$P_j(w)$（w：角周波数）とすると，

$$d_{ij} = \frac{1}{2}\left[\int_{-\pi}^{\pi} 2\left\{\log\frac{P_j(w)}{P_i(w)} + \frac{P_i(w)}{P_j(w)} - 1\right\} dw\right.$$

$$\left. + \int_{-\pi}^{\pi} 2\left\{\log\frac{P_i(w)}{P_j(w)} + \frac{P_j(w)}{P_i(w)} - 1\right\} dw\right]$$

$$= \int_{-\pi}^{\pi}\left\{\frac{P_j(w)}{P_i(w)} + \frac{P_i(w)}{P_j(w)} - 2\right\} dw$$

と定義される。

とし，第 i フレームの自己相関関数を

$$(v_{i0},\ v_{i1},\ \cdots\cdots,\ v_{ip}) \tag{5.7}$$

とすると /x/ との類似度は

$$l(i,\ x) = -\log\left\{\sum_{k=0}^{p} A_{xk}\ v_{ik}\right\} \tag{5.8}$$

で与えられる。

$$\max_{x}\ l(i,\ x) \tag{5.9}$$

を満足する /x/ を第 i フレームの母音認識結果とする。

以上の情報を用いて次の手順で音韻境界を抽出する。

a. 子音の多くは前後の母音に比較して口のせばめ等のためパワーが小さくなる。この性質を利用して音韻境界を抽出する。2種類の閾値 P_T，Q_T を定める。まず，

$$v_{i0} < P_T \tag{5.10}$$

なる区間を音韻境界とする。次に残された区間においてパワーの極小点を求める。極小点において前後の極大点のパワーの差を求め，これを dv_1，dv_2 とする。

$$dv_1 > Q_T \quad \text{かつ} \quad dv_2 > Q_T \tag{5.11}$$

ならその極小点を音韻境界とする。

b. 音韻の境界ではスペクトルが大きく変化する。この性質を使って音韻境界を抽出する。2種類の閾値 E_1，E_2 を定める。まず，

$$e_i^1 > E_1 \tag{5.12}$$

なる区間を音韻境界とする。次に残された区間において極大点を求め，かつ前後の極小点の値の差を求め，これを de_1^1，de_2^1 とする。

$$de_1^1 > E_2 \quad \text{かつ} \quad de_2^1 > E_2 \tag{5.13}$$

ならばその極大点を音韻境界とする。この操作は，系列（5.4）に負符号をつけたものを用いることにより，aとまったく同じ操作で行える。

c. 系列（5.5）にbと同じ操作をほどこして音韻境界を求める。

d. 以上のa，b，cの操作で求められた3種類の音韻境界を統合する。

e. 以上の操作ではまだ良く似た母音間の遷移（たとえば/ao/，/iu/等）の境界の検出が困難である。母音系列を使ってそのような境界を検出する。a～dまでの操作で抽出された音韻境界により分割された区間（以後音韻区間と呼ぶ）のうち継続時間があらかじめ与えられた閾値以上であるものに注目する。母音系列中に同じ母音が連続して4個以上あらわれた場合は，この母音が安定に存在すると考え，これを安定母音と呼ぶ。注目する音韻区間中に安定母音が2個以上存在する場合にはそれらの境界を音韻境界とする。

以上の手順で音韻境界が抽出される。抽出例を図5.3に示す。

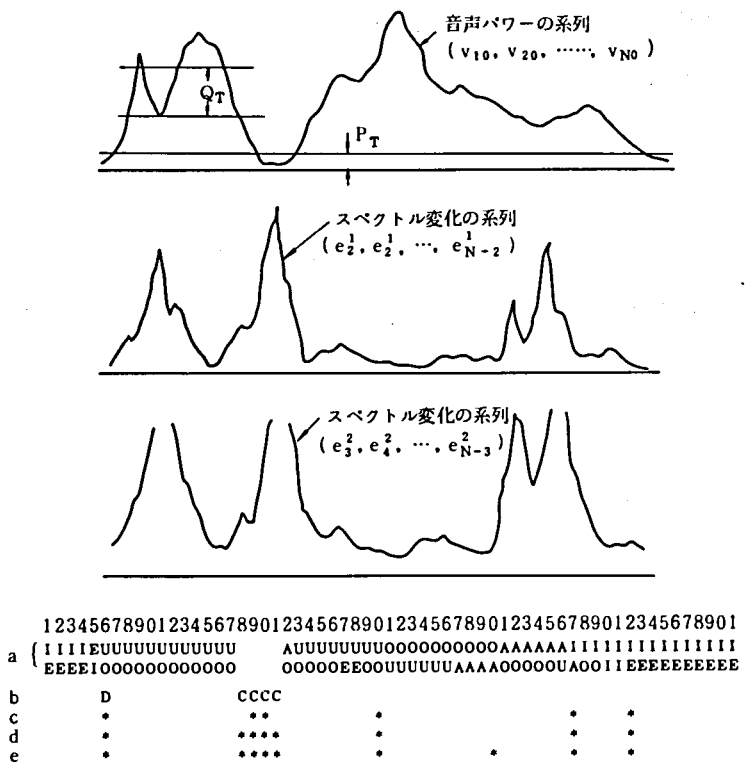


図 5.3　音韻境界抽出手順

a. 第1位，第2位の母音系列
b. 手順aの結果
c. 手順b，cの結果
d. 手順dの結果
e. 手順eの結果（最終結果）

(2) 音韻区間の分類

(1)で決定された音韻区間を母音区間，子音区間に分類する。そのため次の情報を用いる。

a. 各区間の継続時間（区間のフレーム数）

b. 各区間の母音認識結果……区間中の各母音の出現頻度を計算し，最大出現頻度の母音をその区間の母音認識結果とする。

c. 各区間の高域および低域の音声パワー……第 i フレームの 2,000 Hz以上の音声パワーを v_{i0}^{h}，500 Hz 以下の音声パワーを v_{i0}^{l} とすると

$$r^{h} = \frac{\sum_{i \in I} v_{i0}^{h}}{\sum_{i \in I} v_{i0}} \qquad (5.14)$$

（I は注目している区間に属するフレーム番号の集合である）

および

$$r^{l} = \frac{\sum_{i \in I} v_{i0}^{l}}{\sum_{i \in I} v_{i0}} \qquad (5.15)$$

を用いる。

d. 注目している区間の前後の音韻境界の性質

以上の情報と tree 状の判定論理を用いて各区間を母音，子音，未定の 3 種類に分類する。未定という判定をおこなうのは，この段階でまだ確実にはわからない区間を母音，子音のいずれかに強制的に判定することによっておこる誤りを防ぐためである。用いた判定論理を図 5.4 に示す。

母音ないし未定と判定された区間はあらためて母音認識をおこなう。区間の中央 3 フレームを切り出し，それらのフレームの自己相関関数を平均したものをその区間を代表する特徴量として式（5.6）～（5.9）と同様の方法で母音標準パターンとのマッチングにより認識する。ただし，母音の認識誤りを防ぐため候補は 2 個用意する。

(3) VCV音節の抽出

以上の操作で母音ないし未定と判定された区間の中央をセグメント境界とすることによりVCV音節が抽出される。ただし未定と判定された区間を含むセグメントは抽出が一意には定まらない。これは次のように定式化できる。

母音区間の中央のフレームの集合を

$$C^{1} = \{ c_{1}^{1},\ c_{1}^{2},\ \cdots\cdots,\ c_{k_1}^{1} \} \qquad (5.16)$$

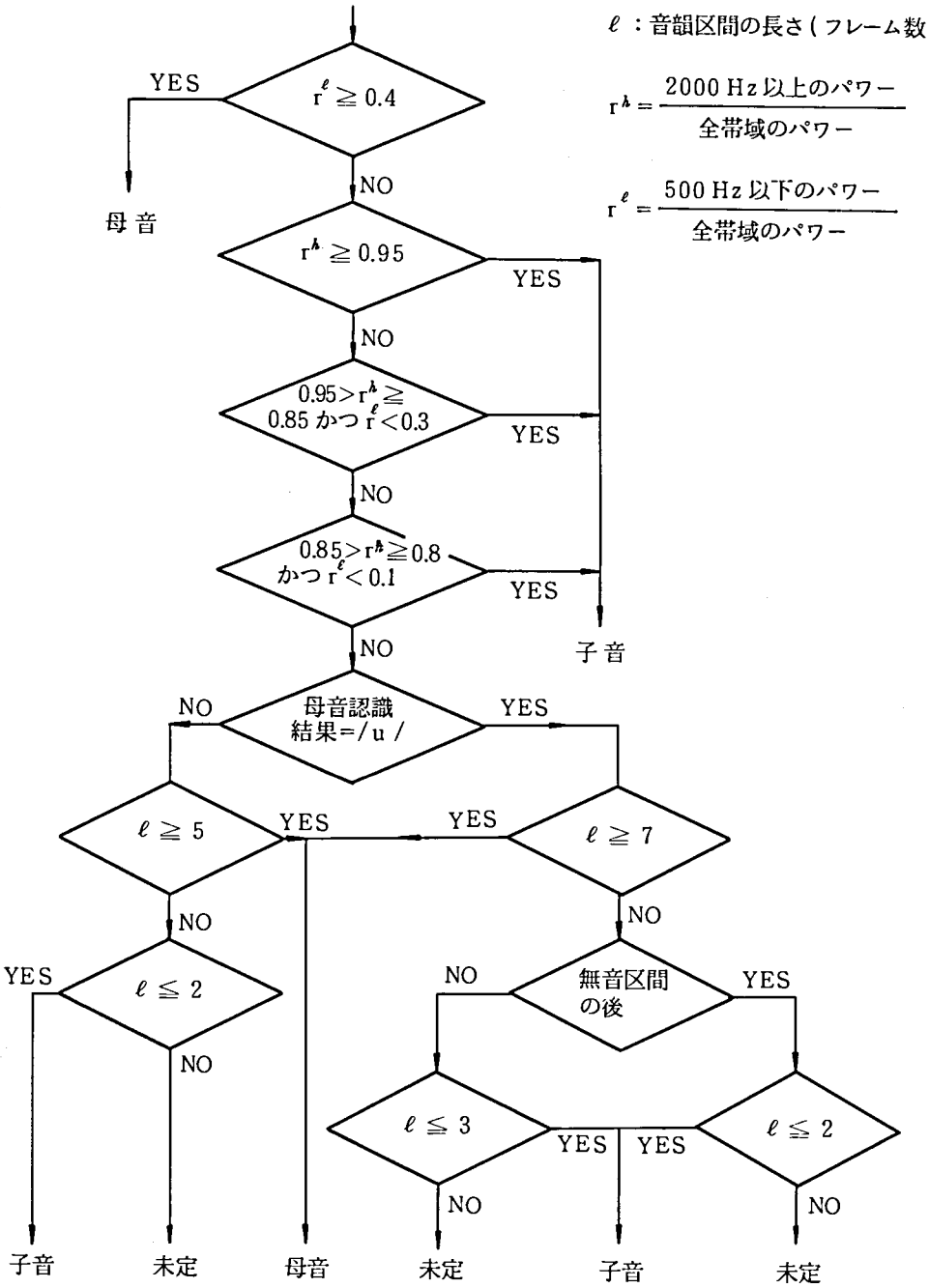


図 5.4　音韻区間分類のための判定論理

とする。同様に未定区間の中央のフレームの集合を，

$$C^2 = \{ c_1^2, c_2^2, \cdots\cdots, c_{k_2}^2 \} \qquad (5.17)$$

とする。C^1，C^2 よりセグメントの境界の集合

$$C = \{ C_1, C_2, \cdots\cdots, C_k \} \qquad (5.18)$$

$$(C_i \in C^1 \cup C^2,\ C^1 \subset C_i)$$

が決まり，VCV音節が抽出される。このようにして抽出されるすべての異なったＶＣＶ音節が候補となる。図 5.3 に示したのと同じ例についてVCV音節を抽出した結果を図5.5 に示す。以上の結果をまとめて，セグメント認識部へは次の情報が送られる。

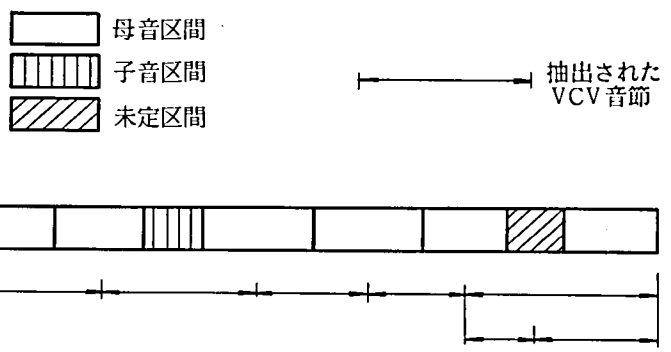


図 5.5　VCV音節の抽出

a.　相関関数の時系列

b.　各VCV音節の先頭と末尾のフレーム番号

c.　各VCV音節の先頭と末尾の母音の候補

図 5.3，図 5.5 と同じ例について，認識部へ送られる情報を図 5.6 に示す。

5.3.3　セグメント認識部

セグメント認識部では，入力音声から切り出されたVCV音節とVCV音節標準パターンとのDPを用いた時間正規化マッチングにより各セグメントの認識をおこなう。手順を次に示す。

(1)　相関関数の時系列として表現されている音声区間からVCV音節に相当する部分を切り出す。切り出されたVCV音節の 1 つを，

$$(v_1, v_2, \cdots\cdots, v_{N_1}) \qquad (5.19)$$

$$(\boldsymbol{v}_i = (v_{i0}, v_{i1}, \cdots\cdots, v_{ip}))$$

とする。

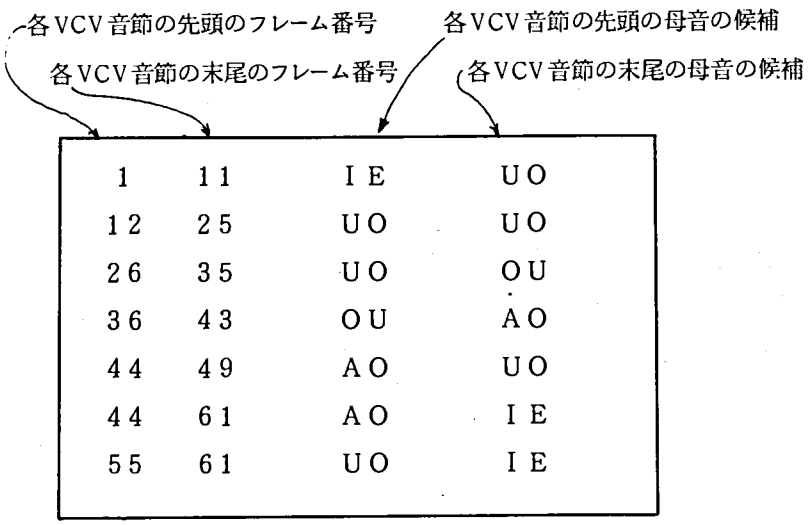


図 5.6　認識部へ送られる情報

(2)　VCV音節標準パターンは最尤スペクトルパラメータの時系列としてたくわえられている。切り出されたVCV音節の母音部分はすでに候補が決まっているから，母音部分の一致する標準パターンとのみマッチングをおこなう。マッチングすべき標準パターンの1つを

$$(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots\cdots, \boldsymbol{A}_{M1}) \qquad (5.20)$$

$$(\boldsymbol{A}_j = (A_{j0}, A_{j1}, \cdots\cdots, A_{ip}))$$

とする。

(3)　式(5.19)，式(5.20)から類似度マトリクスLMを作成する。

$$LM = \{ l(i, j) \} \qquad (5.21)$$

$$(i = 1, 2, \cdots\cdots, N_1, \quad j = 1, 2, \cdots\cdots, M_1)$$

ただし，$l(i, j)$は入力のVCV音節の第iフレームと標準パターンの第jフレームの類似度で，

$$l(i,\ j) = -\log\left\{\sum_{k=0}^{p} A_{jk}\ v_{ik}\right\} \qquad (5.22)$$

で与えられる。

(4) 類似度マトリクス上で類似度和が最大のパスを探索する。すなわち，

$$L = \max_{f}\left\{\sum_{i=1}^{N_1} l(i,\ f(i))\right\} \qquad (5.23)$$

を満足するLを求める。ただし，関数fは適度な時間軸の正規化が行われるように制限を加える必要がある。ここでは，次のような3つの自由度を許すこととした。

$$f(i) = \begin{cases} f\ (i-1) \\ f\ (i-1)+1 \\ f\ (i-1)+2 \end{cases} \qquad (5.24)$$

$$(f(1) = 1,\ f(N_1) = M_1,\ i = 2,\ 3,\ \cdots\cdots,\ N_1)$$

式（5.24）の条件のもとで式（5.23）を満足するLを求めると，これが式(5.19)，(5.20)の間で時間軸の正規化をおこなって得られた類似度である。この様子を図5.7に示す。この計算はDPを用いて容易におこなうことができる。

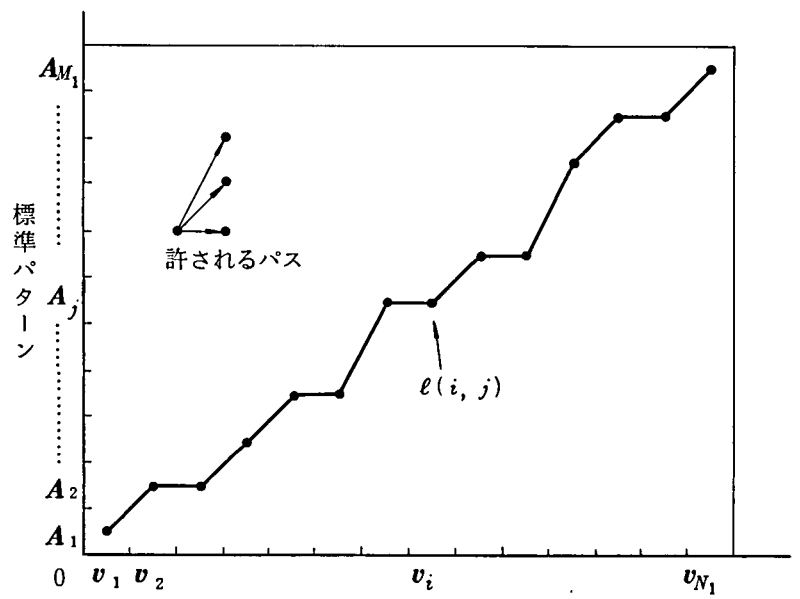
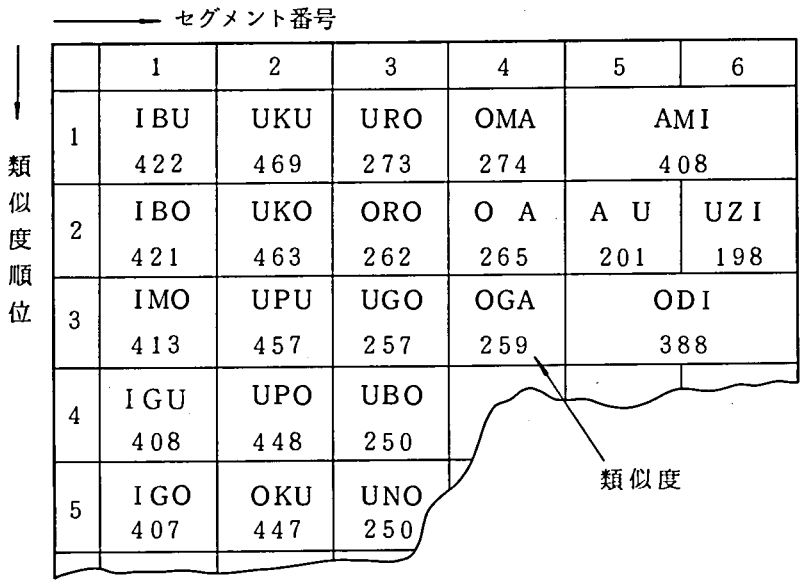


図5.7　セグメントの認識におけるDPマッチング

(5) (2)で述べた条件を満足するすべての標準パターンとのマッチングをおこない，類似度を求める。

認識結果は類似度第1位のもののみを採用してもよいが，それでは情報の損失となり好ましくない。したがって，認識結果は類似度と共にすべて次段へ送ることにする。すなわち，セグメント化部の出力は類似度順位，および各セグメントの時間順序に従って整理され次段へ送られる。図5.3等と同じ例について，VCVマトリクスを図5.8に示す。



	1	2	3	4	5	6
1	IBU 422	UKU 469	URO 273	OMA 274	AMI 408	
2	IBO 421	UKO 463	ORO 262	O A 265	A U 201	UZI 198
3	IMO 413	UPU 457	UGO 257	OGA 259	ODI 388	
4	IGU 408	UPO 448	UBO 250			
5	IGO 407	OKU 447	UNO 250			

図5.8　VCVマトリクス

5.4 連続単語認識部 (70)(24)(119)

連続単語認識部では，音響処理部から送られてくるVCVマトリクスから最もたしからしい単語系列を求める働きをする。5.2節で述べたようにその基本となる考え方は，すべての単語系列を作り出し，各単語系列ごとに入力との類似度を求め，最大の類似度をもつ単語系列を選ぶという手法である。したがって連続単語認識部は基本的には次の2つの機能をもった部分から構成される。

(1) 任意の単語系列と入力音声との類似度を求める機能

(2) 最大の類似度をもつ単語系列を選択する機能

この処理をおこなうには膨大な計算量が必要になるため，実際にはDPを用いて計算量の減

少をはかる。DPは前の処理結果を使いながら1ステップづつ計算を進めてゆく手法である。したがって，DPを用いれば単語系列ごとに類似度を求める必要はなく，単語単位で類似度を求める機能をもっていればよい。つまり上の2つの機能は次のように書きなおすことができる。

(1) 入力音声の任意の部分と単語との類似度を求める機能

(2) DPにより，最大類似度をもつ単語系列を求める機能

ここで用いた連続単語認識部では単語認識部と単語系列処理部が上の機能をはたすことになる。そのほかに単語認識の際用いる単語辞書があり，計3つの部分から構成されている。各部のくわしい動作を次に説明する。

5.4.1 単語辞書

単語辞書中では各単語がVCV音節の系列としてたくわえられている。たとえば / ehagaki /（絵葉書）という単語は単語辞書中で次のように表現される。

$$\text{ehagaki} \rightarrow \underline{\text{eha}} \quad \underline{\text{aga}} \quad \underline{\text{aki}} \tag{5.25}$$

認識対象の変更，増加は単語辞書の内容をかきかえることにより，容易におこなうことができる。

5.4.2 単語認識部

VCVマトリクス上の任意の部分が指定されたとき，そこにおける各単語の類似度を求める機能をもつ。いま単語の始点，終点がVCVマトリクス上の第i列，第j列（$i \leqq j$）として指定されたとする。単語wを構成しているVCV音節をvcv^1, vcv^2, …… vcv^kとし，（i, j）区間におけるwとの類似度を$L(i, j \mid w)$とする。区間内に存在するVCV音節のセグメント数がk個であり，各セグメントにおけるvcv^1, vcv^2, ……, vcv^kとの類似度が

$$L(vcv^1),\ L(vcv^2),\ \cdots\cdots,\ L(vcv^k) \tag{5.26}$$

であるとすると

$$L(i, j \mid w) = \sum_{i=1}^{k} L(vcv^i) \tag{5.27}$$

とする。もしk個のセグメントのとり方が一意に定まらない場合にはすべてのとり方について式（5.26）を計算し，その最大値をあらためて$L(i, j \mid w)$とする。また，（i, j）区間内

に存在するセグメントの数がk個でない場合は

$$L(i, j \mid w) = 0 \qquad (5.28)$$

とする。すべての認識対象の中で最大の類似度を持つ単語，すなわち

$$L(i, j) = \max_{w} L(i, j \mid w) \qquad (5.29)$$

を満足する単語$w(i, j)$を認識結果とし，そのときの類似度$L(i, j)$と共に単語系列処理部へ送る。

5.4.3 単語系列処理部

単語認識部から送られてくる情報をもとに，音声区間全体での類似度和を最大にするという評価基準のもとで最適の単語系列を求める。すなわち，

$$L_m = \max_{1 \leqq i_1 < i_2 \cdots < i_n < m} \left\{ L(1, i_1) + L(i_1+1, i_2) + \cdots\cdots + L(i_n+1, m) \right\} \qquad (5.30)$$

（mはVCVマトリクスの列の数）

を満足する単語系列

$$w(1, i_1)\ w(i_1+1, i_2)\ \cdots\cdots\ w(i_n+1, m) \qquad (5.31)$$

が連続単語の認識結果であるとする。式（5.30）をそのまま計算するのは大変であるが，これは次のように漸化式表現になおすことができる。

$$L_0 = 0, \quad L_j = \max_{i=1}^{j} \left\{ L_{i-1} + L(i, j) \right\} \qquad (5.32)$$

これはDPの適用が可能なことを示しており，処理量を大幅に減らすことができる。図5.8に示したVCVマトリクス上でDPを用いて実際に処理を進めてゆく手順を図5.9に示す。

式（5.32）をみるとm（すなわち音声区間の長さ）の2乗に比例して処理量が増すようになっているが，実際には単語長には上限があるため式（5.32）は次のように書きなおすことができる。

$$L_0 = 0, \quad L_j = \max_{i=j-l_w}^{j} \left\{ L_{i-1} + L(i, j) \right\} \qquad (5.33)$$

（l_wは最長の単語のセグメント数）

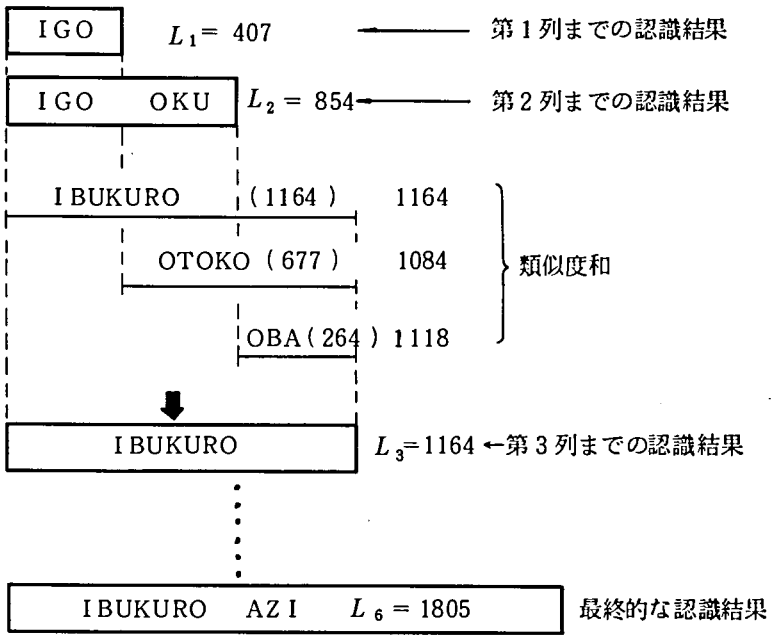


図 5.9　連続単語認識の手順

すなわち，図 5.10に示したように，L_jを求めるには直前のl_w個のL_iのみを使って計算すればよいことになる。これは，処理量が音声の継続時間に比例して単に線形に増大するのみであることを示している。したがってこの方法は入力音声の先頭から順に処理してゆくことができ

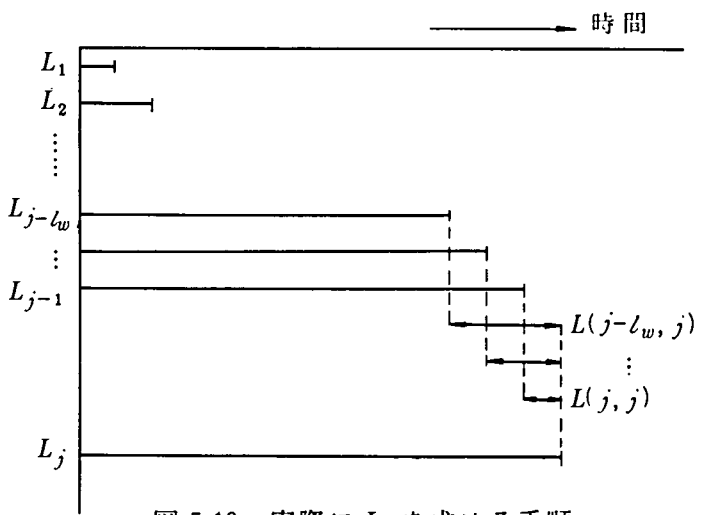


図 5.10　実際にL_jを求める手順

（L_jを求めるには，L_{j-l_w}，L_{j-l_w+1}，……，L_{j-1}を用いれば良いことを示している）

しかも音声区間全体での最適解が得られるため，きわめて実際上便利な方法である。

5.5 認識実験と検討

5.5.1 認識対象

連続単語音声を構成している単語は表5.1に示した90種類の単語である。

表5.1 認識対象の単語

VCV型 uta（歌），oba（おば），eki（駅），imu（医務），ato（後），oni（鬼），iro（色），one（尾根），aze（あぜ），isu（いす），ichi（一），oku（奥），ura（裏），ase（汗），ima（居間），esa（えさ），igo（囲碁），ada（仇），ushi（牛），oto（音），eri（えり），ebi（海老），ehu（F），ume（梅），uha（右派），azi（味），edo（江戸），ike（池），ono（オノ），igi（異議），

VCVCV型 omote（表），ihuku（衣 服），akebi（あけび），abura（油），umibe（海 辺），otsuge（お 告），odosu（威 す），onushi（お 主），agari（上 り），aniki（兄 貴），akago（赤 子），eboshi（えぼし），umare（生 れ），ichiza（一 座），ikusa（戦），ahiru（アヒル），otaku（お 宅），enogu（絵 具），udegi（腕 木），asobi（遊 び），irori（いろり），otoko（男），ibara（茨），opera（オペラ），umezu（梅 酢），adana（あだ名），asemo（あせも），ohana（お花） eziki（餌 食），uneri（うねり），

VCVCVCV型 amehuri（雨降り），azakeri（喋り），anaguma（穴熊），ibaragi（茨木），inemuri（居眠），udegumi（腕組み），umeboshi（梅干し），esupuri（エスプリ），ogakuzu（おがくず），odoriko（踊り子），ohitashi（おひたし），ezomatsu（えぞ松），ibukuro（胃 袋），azemichi（あぜ道），ichiziku（いちじく），ehagaki（絵葉書），ebigani（えびがに），usemono（うせ物），otedama（お手玉），otohime（乙 姫），akagire（あかぎれ），asakaze（朝 風），agezoko（あげ底），agohige（あごひげ），arupaka（アルパカ），inutade（いぬたで），usotsuki（うそつき），abekobe（あべこべ），unohana（うの花），uketsuke（受 付），

そのうちわけを次に示す。

VCV型単語 ……30種類

VCVCV型単語 ……30種類

VCVCVCV型単語……30種類

これらの単語を2～4個連続して発声した90種類の単語を認識対象とする。そのうちわけは次のとおりである。

単語を2個連続して発声したもの……30種類

〃　3個　　〃　　30種類

〃　4個　　〃　　30種類

これらの連続単語は乱数を用いて90種類の単語からランダムに選んだ単語をつなぎあわせることにより作成した。連続単語音声サンプルのリストを表5.2に示す。これらのサンプルを3名の発声者（発声者T，I，Sと呼ぶ）が各2回発声した計540サンプルを認識実験に用いた。これらのサンプル中には計1,620個の単語が含まれている。なお各サンプルを発声する際には単語中での無声化は許すこととし，語尾での無声化は許さないこととした。

表5.2　連続単語のリスト

1.　2個

オノ・茨木，	乙姫・あかぎれ，	アルパカ・茨
鬼・あべこべ，	表・えり，	御主・えぼし
えぞ松・餌食，	江戸・上り，	あげ底・居眠
F・海辺，	あぜ・お告，	遊び・いぬたで
胃袋・味，	うそつき・威す，	えり・おひたし
うねり・表，	雨降り・異議，	赤子・茨木
嘲り・茨，	上り・あぜ，	えさ・腕組み
あせも・生れ，	江戸・裏，	えび・威す
絵具・いす，	赤子・おひたし，	音・一
牛・囲碁，	おがくず・F，	穴熊・あべこべ

表 5.2　（続　き）

2.　3 個

歌・アヒル・あかぎれ，えび・絵具・乙姫，梅酢・えびがに・一

異議・汗・居眠，オノ・いす・腕木，遊び・医務・おひたし

油・奥・アルパカ，えぞ松・梅干し・囲碁，御主・いぬたで・居間

池・鬼・上り，駅・オノ・医務，あごひげ・赤子・音

茨・遊び・奥，茨木・あぜ道・絵葉書，あぜ・エスプリ・あけび

胃袋・右派・居眠，油・えぞ松・梅酢，喇り・遊び・餌食

穴熊・あごひげ・お告，乙姫・池・お告，牛・味・遊び

えぞ松・梅酢・乙姫，オペラ・梅酢・エスプリ，お花・御主・奥

上り・あべこべ・後，赤子・いぬたで・オノ，威す・仇・アルパカ

色・鬼・うせ物，衣服・うの花・音，うねり・えぼし・おひたし

3.　4 個

踊り子・おば・海辺・えぞ松，味・遊び・あだ名・居間

いちじく・音・衣服・お花，歌・腕木・居間・味

奥・赤子・おば・あぜ道，喇り・朝風・アルパカ・乙姫

オノ・あげ底・胃袋・絵具，江戸・雨降り・上り・あぜ

御主・居間・おひたし・あぜ道，梅酢・穴熊・えびがに・あけび

赤子・上り・裏・一座，お手玉・うせ物・腕組み・汗

味・喇り・F・奥，絵具・池・あぜ・江戸

居眠・エスプリ・お宅・アルパカ，異議・あだ名・右派・威す

えぼし・茨・うせ物・赤子，生れ・居間・梅干し・表

一・アヒル・喇り・男，居間・いす・囲碁・男

遊び・アヒル・生れ・踊り子，駅・生れ・遊び・いす

衣服・あぜ・お手玉・いちじく，遊び・腕組み・乙姫・男

うせ物・鬼・あぜ道・生れ，乙姫・居眠・えぞ松・上り

茨木・あかぎれ・お告・お宅，油・あせも・雨降り・うせ物

上り・受付・居間・奥，一・おひたし・牛・梅

5.5.2 実際上の問題点とその対策

本連続単語認識系は音響処理と系列処理という2段がまえの処理をし，しかも各処理部においてDPが主要な役割をはたしているという，簡潔な構成になっている。しかしながら実際の認識をおこなうにあたってはいくつかの実際的な考慮をはらう必要がある。これらについて次に述べる。

(1) 単語境界

5.4.2で述べたように単語辞書中では各単語はVCV音節の系列としてたくわえられている。しかしながら図5.8のVCVマトリクスの例を見るとわかるように，単語境界においては前の単語の語尾の母音と次の単語の最初の母音が合同して1つのVCV音節（この場合はVV型の音節）が余分にあらわれる。類似度和を計算するにはこの音節の類似度も考慮に入れるべきである。この問題の対策として，ここでは単語辞書中で各単語ごとに次に示すような2種類の表現をたくわえておくこととした。

$$\text{ehagaki（絵葉書）} \rightarrow \begin{cases} \text{(a)} & \underline{\text{eha}} \quad \underline{\text{aga}} \quad \underline{\text{aki}} \\ \text{(b)} & \underline{*\ \text{e}} \quad \underline{\text{eha}} \quad \underline{\text{aga}} \quad \underline{\text{aki}} \end{cases} \qquad (5.34)$$

ただし，＊はどの音韻であっても良いことを示す。実際上は＊は前の単語の最後の母音にするべきであるが，DPの計算の際類似度和以外に認識結果も記憶しておく必要がある等の問題が生じるので(b)のように表現するにとどめた。(a)の表現もたくわえておいたのは次のような場合に必要だからである。

(1) 連続単語の語頭

(2) 下に示す例のように単語の最後の母音と次の単語の先頭の母音が一致する場合（この場合にはVV型のセグメントは生じない）

aze ehagaki

辞書の内容を引く場合には，各場合ごとに(a)，(b)いずれの表現が正しいかを判定し，正しい方の表現を用いるべきであるが，ここでは簡単化のため語頭を除いていずれの表現も同等に扱うこととした。なお単語辞書の変更の際には式（5.34）の左辺の形でデータを入力すれば自動的に右辺の表現に変換されてたくわえられる構成になっている

(2) 無声化の対策

現実の音声データでは無声化が生じることはさけられない。本認識実験でも語中の無声化を許している。したがって認識系は無声化に関する規則をあらかじめ知っていてこれを利用して正しい認識をおこなう必要がある。無声化の規則は一種の音韻変形規則としてあらわされる。入力の音韻系列に規則を適用して正しい音韻系列に復元する方法（bottom-up的手法）と，単語辞書の内容に規則を適用して正しい書式を実際の入力音韻系列に変換する方法（top-down的手法）が考えられるが，ここでは後者の方法をとった。このようにすれば無声化した場合の音韻系列をあらかじめ辞書に登録しておくという方法でよく，認識部の構成が複雑になるのをさけることができる。あらかじめいくつかの予備サンプルを用意して無声化のおこりやすい単語を抽出し，そのような単語について無声化した場合に得られるVCV音節の系列を辞書の内容に用意しておくことにした。表5.3に無声化する単語についての辞書の内容を示す。

表5.3　無声化を考慮した辞書の内容

正しいつづり	無声化を考慮したつづり	辞書の内容
usotsuki	usoki	uso oki (* u uso oki
ohitasi	osasi	osa asi (* o osa asi
esupuri	esuri	esu uri (* e esu uri
uketsuke	ukete	uke ete (* u uke ete
usi	us *	us * (* u us *
ichi	ich *	ich * (* i ich *
onusi	onus *	onu us * (* o onu us *
ebosi	ebos *	ebo os * (* e ebo os *
ohitasi	ohitas *	ohi ita as * (* o ohi ita as *

(3) 音韻間距離行列

図 5.8 に示したVCVマトリクスは記憶容量の制限等のために実際には適当な行数で切る必要がある。この場合，VCVマトリクスにのってないVCV音節の類似度は適当な方法で推定する必要がある。この推定を次のような方法でおこなう。

$u_1\ u_2\ u_3$ なるVCV音節の類似度を $L(u_1\ u_2\ u_3)$ とする。このとき $u_1'\ u_2'\ u_3'$ なるVCV音節の類似度 $L(u_1'\ u_2'\ u_3')$ を次の式で推定するものとする。

$$L(u_1'\ u_2'\ u_3') = L(u_1\ u_2\ u_3) - \sum_{i=1}^{3} d(u_i\ u_i') \qquad (5.35)$$

ただし $d(u_i\ u_i')$ は音韻 u_i と u_i' の距離を示す値である。式（5.35）を計算するためには各音韻間の d が与えられている必要がある。ここでは経験的に，表 5.4 に示す音韻間距離行列を用いた。音韻 /x/ の行，音韻 /y/ の列の要素が音韻間の距離 d（x, y）を示している。＊はいずれの音韻でもよいという仮定であったから表 5.4 に示すようにいずれの音韻も＊との距離は 0 になっている。また $L(u_1\ u_2\ u_3)$ は類似度第 1 位のVCV音節の値を用いた。

5.5.3 連続単語音声認識結果

連続単語中の単語の認識結果を表 5.5 に示す。全体で 92.2 %の認識結果が得られた。次にこの結果をくわしく見てみる。

(1) 単語の位置と誤りの関係

単語が連続単語中に占める位置によって認識率がどのように変化するかを調べるため，単語の位置を語頭，語中，語尾の 3 種類に分けて結果を分類した。その結果を表 5.6，図 5.11に示す。発声者間に少々差が見られるが，全体として語頭，語尾の単語は同じ程度の誤りが生じており，また語中の単語はそれらに比較してほぼ 2 倍近い誤りが生じているということがいえる。語中の単語に誤りが生じやすいのは単語の始点，終点が共にあらかじめわかっていないためであり直観的に明らかである。これに対し語尾の単語が語頭と同じ程度，確実に認識できることは注目される。これはここで用いた認識方法が，音声全体をみて最適な単語系列を決定するというアルゴリズムになっているためである。迫江，千葉の方法[34]は音声の先頭から順に認識していくため，誤りが累積してゆき語尾の単語は語頭に比較して認識率が低下することが予想され，その点において本方法の方が有利である。

孤立単語を含めて考えると，単語は境界条件によって次の 3 種類に分類される。

表 5.4　音韻間距離行列

母音

	a	i	u	e	o	*
a	0	50	35	35	25	0
i	50	0	20	25	40	0
u	35	20	0	20	20	0
e	35	25	20	0	25	0
o	25	40	20	25	0	0
*	0	0	0	0	0	0

子音

	m	n	b	d	g	r	z	p	t	k	s	h	•	*
m	0	5	10	10	10	10	10	25	25	25	30	30	20	0
n	5	0	10	10	10	10	10	25	25	25	30	30	20	0
b	10	10	0	5	5	10	10	10	10	10	20	20	20	0
d	10	10	5	0	5	10	10	10	10	10	20	20	20	0
g	10	10	5	5	0	10	10	10	10	10	20	20	20	0
r	10	10	10	10	10	0	10	20	20	20	15	15	10	0
z	10	10	10	10	10	10	0	20	20	20	15	15	20	0
p	25	25	10	10	10	20	20	0	5	5	10	10	40	0
t	25	25	10	10	10	20	20	5	0	5	10	10	40	0
k	25	25	10	10	10	20	20	5	5	0	10	10	40	0
s	30	30	20	20	20	15	15	10	10	10	0	5	40	0
h	30	30	20	20	20	15	15	10	10	10	5	0	40	0
•	20	20	20	20	20	20	10	40	40	40	40	40	0	0
*	0	0	0	0	0	0	0	0	0	0	0	0	0	0

表 5.5　連続単語認識結果

対象 / 発声者	2 連続単語	3 連続単語	4 連続単語	計
T	114/120 95.0 %	161/180 89.4 %	215/240 89.6 %	490/540 90.7 %
I	109/120 90.8 %	164/180 91.1 %	224/240 93.3 %	497/540 92.0 %
S	110/120 91.7 %	173/180 96.1 %	223/240 92.9 %	506/540 93.7 %
計	333/360 92.5 %	498/540 92.2 %	662/720 91.9 %	1,493/1,620 92.2 %

表 5.6　単語の位置と誤りの関係

単語の位置 / 発声者	語　頭	語　中	語　尾	計
T	11 (22.0 %)	28 (56.0 %)	11 (22.0 %)	50 (100 %)
I	10 (23.3 %)	21 (48.8 %)	12 (27.9 %)	43 (100 %)
S	11 (32.4 %)	11 (32.4 %)	12 (35.3 %)	34 (100 %)
計	32 (25.2 %)	60 (47.2 %)	35 (27.6 %)	127 (100 %)

a．両端のが既知のもの（孤立単語）

b．一端のみ既知のもの（連続単語の語頭，語尾の単語）

c．両端とも未知のもの（連続単語の語中の単語）

各単語の認識率を求め図 5.12に示す。ただし孤立単語の認識結果は第 4 章の同じ 3 名の発声者のものを用いた。3 種類の認識率がほぼ直線上に並んでいる。理論的根拠は明らかでないが興味深い現象である。

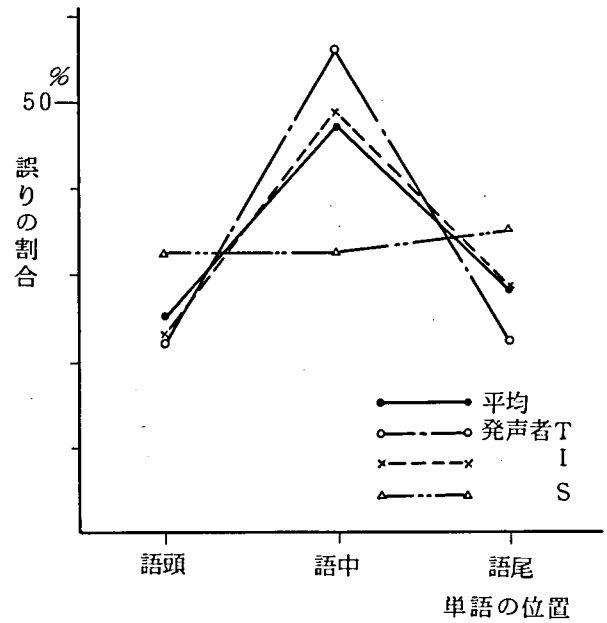


図 5.11　単語の位置別に分類した誤りの場合

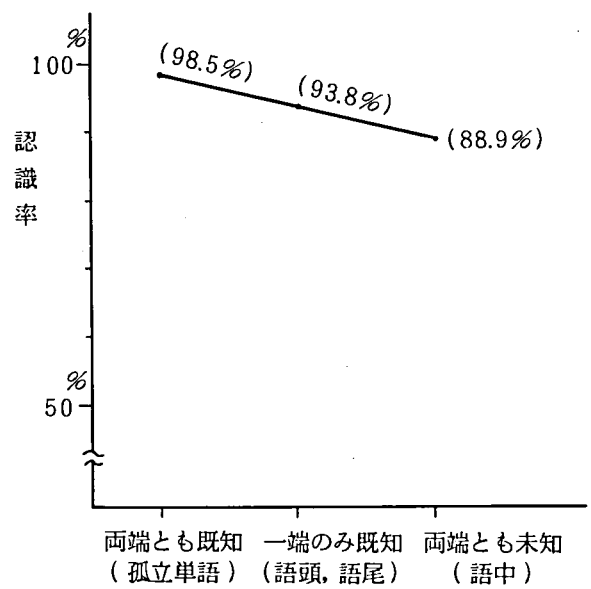


図 5.12　単語境界の情報と認識率の関係

(2)　単語数と認識率の関係

連続した単語の数が増えると認識率がどのように変化するかを見るために連続単語を構成している単語数と認識率の関係を図 5.13に示す。比較のため図 5.12と同様に同じ 3 名についての単語認識結果を示してある。孤立単語から 2 連続単語への認識率の低下は大きいが，それ以後は単語数が増加しても認識率の低下はわずかである。したがってより長い連続単語音声も安定

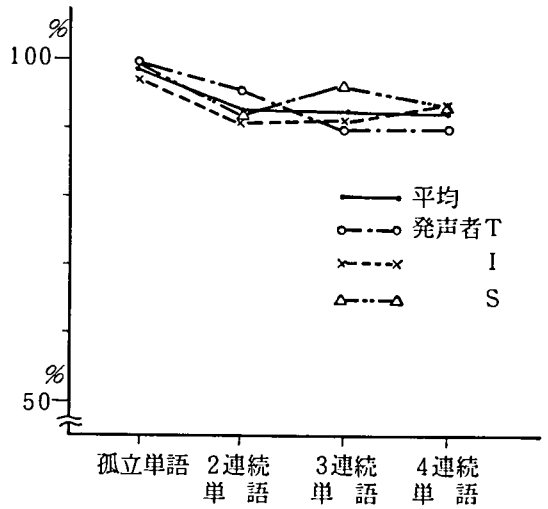


図 5.13　単語数と認識率の関係

した認識がおこないうると期待される。この裏づけとして連続単語認識のモデルを考えてみる。すなわち(1)の考察の結果から連続単語中では語頭，語尾の単語および語中の単語の認識率はそれぞれ一定で，次のように定まると仮定する。

語頭，語尾の単語の認識率……P_1

語中の単語の認識率 ……P_2

するとn個の単語が連続したn連続単語の認識率は次式で与えられる。

$$\frac{2P_1 + (n-2)P_2}{n} \quad (n \geqq 2) \tag{5.36}$$

P_1，P_2は(1)の結果より次のように定まる。

$$P_1 = 93.8\%, \qquad P_2 = 88.9\% \tag{5.37}$$

このモデルによる理論値と実際の値の比較を図14に示す。両者はかなり良く一致していることがわかる。2連続単語では若干のずれが認められるがこれは次の理由によると考えられる。

a. モデルが適用できる限界の場合であるから，式（5.36）のような簡単な式と実際の値にある程度の誤差がでるのはやむをえない。

b. 短い連続音声ではセグメント化の誤りが致命的に認識結果に響きやすい。(逆に長い連続音声では，セグメント化の際，挿入誤りと脱落誤りが生じた場合，全体としてのセグメント数は正しい値となるためセグメント化の段階で回復できる可能性がある。したが

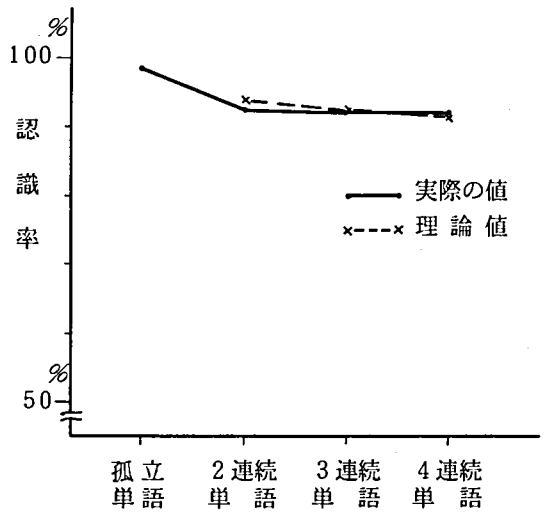


図 5.14　理論値との比較

ってセグメント化の誤りが認識結果に及ぼす影響はより少ない。）

以上のような考察から，式（5.36）のような簡単なモデル式がかなり妥当性をもつといえる。なお式（5.36）でnを無限大にすると認識率はP_2に収束する。すなわち，連続する単語数が多くなっても安定に認識できることを示している。

(3)　誤りのタイプ

認識誤りは次の3つのタイプに分けることができる。

a.　1つの単語が複数個の単語に分裂して認識される場合（分裂）

b.　別の単語に誤認識される場合（置換）

c.　2つ以上の単語が合同して1つの単語として認識される場合（合同）

各タイプの間の誤りの割合を調べるため，誤りをタイプ別に分類して表5.7，図5.15に示す。分裂，置換の誤りはほぼ同じ程度生じており，これに対し合同の誤りは非常に少ないことがわかる。これは定性的には次のように説明できる。連続単語認識のアルゴリズムは図5.16のようにモデル化できる。すなわち，VCV型，VCVCV型，VCVCVCV型各単語に対応した3種類の写像関数f_1，f_2，f_3が入力音声を単語空間に写像するものとする。誤認識が生じる場合，各単語が生じる割合は単語空間における3種類の単語の密度に比例すると考えられる。

表 5.7　タイプ別の誤りの分類

発声者＼タイプ	a（分裂）	b（置換）	c（合同）	計
T	18 (36.0%)	30 (60.0%)	2 (4.0%)	50 (100%)
I	20 (46.5%)	18 (41.9%)	5 (11.6%)	43 (100%)
S	17 (50.0%)	17 (50.0%)	0 (0%)	34 (100%)
計	55 (43.3%)	65 (51.2%)	7 (5.5%)	127 (100%)

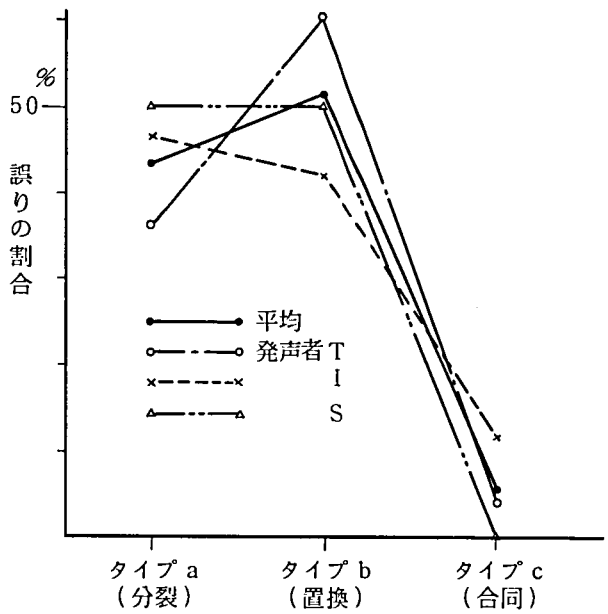


図 5.15　タイプ別に分類した誤りの割合

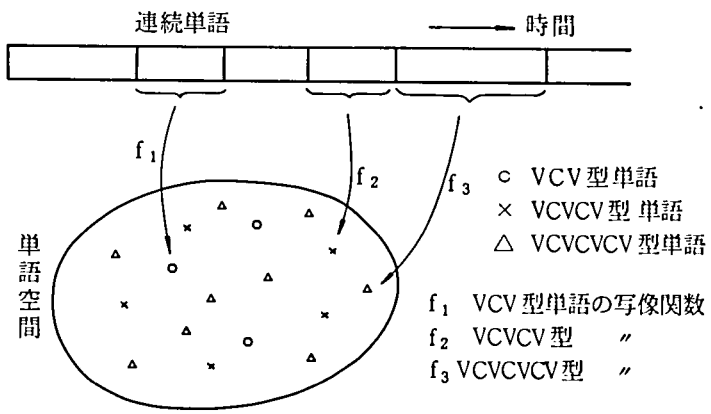


図 5.16　連続単語認識のモデル

誤認識された結果がVCV型単語である割合…… $\frac{k}{n_v^2 n_c}$

誤認識された結果がVCVCV型単語である割合…… $\frac{k}{n_v^3 n_c^2}$

誤認識された結果がVCVCVCV型単語である場合…… $\frac{k}{n_v^4 n_c^3}$

（5.38）

ただし、k：定数，n_v：母音数（＝5），n_c：子音数（＝12）

また分裂，置換，合同により生じる単語の種類は次のようになる。

分裂……VCV型，VCVCV型

置換……VCV型，VCVCV型，VCVCVCV型

合同……VCVCV型，VCVCVCV型

（5.39）

（5.38），（5.39）より各誤りの生じる割合はほぼ次のようになる。

分裂：置換：合同≒VCV型単語の生じる割合：VCV型単語の生じる割合：ＶＣＶＣＶ型単語の生じる割合≒$n_v n_c$：$n_v n_c$：1＝60：60：1　　（5.40）

式（5.40）により分裂，置換の生じる割合が同じ程度であること，また合同の生じる割合はこれらよりずっと少ないことが説明された。

5.5.4 誤りの分析

認識誤りの原因はセグメント化の誤りによるものとVCV音節の認識誤りによるものとがある。誤りを原因別に分類して表5.8に示す。次の各誤りについてくわしく分析する。

(1) セグメント化の誤り

セグメント化の段階での誤りが原因となったものである。これは更にa.音韻境界抽出の誤り，b.音韻区間の分類誤りに分けられる。前者はa.挿入誤り，b.母音境界の抽出できなかった誤り，c.子音の検出できなかった誤り，が同じ程度の割合になっている。後者はa.母音区間を子音と判定した誤り，b.子音区間を母音と判定した誤りに分けられる。特に子音区間を母音と判定し

表 5.8　誤りの原因別の分類

誤りの原因 ＼ 発声者	セグメント化					セグメントの認識	
	音韻境界抽出			音韻区間分類		母音認識	VCV音節の認識
	脱落		挿入	母音→子音	子音→母音		
	母音境界	子音検出					
T	3	4	0	5	7	10	21
I	1	3	3	1	4	4	27
S	0	1	3	0	9	7	14
計	4	8	6	6	20	21	62
	44（34.6％）					83（65.4％）	

た誤りが多く，これだけでセメント化の誤りの 45.5％を占めている。したがってセグメント化の性能を上げるためにはこの誤りを少なくする必要がある。そこでこの誤りを子音別に分類して表 5.9 に示す。表 5.9 より誤りのほとんどは鼻音および摩擦音の場合に生じていることがわかる。これらの子音では子音区間が長く，しかも定常的であるため分類が困難と考えられる。これらの誤りを少なくするには次のような対策をたてる必要がある。

表 5.9　音韻区間分類誤りの分析
（子音→母音）

子　音	m	n	g	z	h	s	母音境界	計
個　数	1	3	5	5	2	3	1	20

鼻　音……標準パターンを用意して認識をおこない母音と区別する。

摩擦音……スペクトルの局所的な情報（すなわち高域に多くの成分をもつという情報）の使い方を精密化して母音と区別する。

全体としてセグメント化による誤りは全体の 34.6％で VCV 音節の認識による誤りのほぼ半分である。これは後で述べるように無声化による誤りを辞書の内容の変更により救うことができ

たことが大きな原因である。

(2) VCV音節の認識誤り

VCV音節の認識が正確におこなえなかったことが原因で生じた誤りである。そのうち25.3%は母音認識が誤ったため正しいVCV音節がVCVマトリクス中にあらわれなかったものである。現在母音の候補は2つとっているが，この結果を見るとまだかなり母音の誤りが生じていることがわかる。どのような状況で母音の誤りが生じているかを調べると，a.鼻音，b.無声摩擦音，c.無声破裂音のいずれかの子音の前後で生じていることがわかる。これらの統計をとったものを表5.10(a)に示す。鼻音の前後，無声摩擦音の後，無声破裂音の前の母音が誤りやすいことがわかる。表5.10(b)はこれらの子音の組み合わせによる母音の誤りの統計を示す。これら3種類の子音の組み合わせにはさまれた母音が誤りやすいことを示している。この対策としては，変形した母音の標準パターンを別に用意しておき，子音認識をおこなった後変形しやすい位置にある母音を認識しなおすという操作が必要であろう。

表5.10 母音認識誤りの生じている状況

a.

位置	鼻音		無声摩擦音		無声破裂音	
	前	後	前	後	前	後
個数	9	9	2	6	6	0

b.

位置	単独	鼻—鼻	鼻—破	摩—鼻	摩—破	鼻—摩	計
個数	10	3	3	3	1	1	21

鼻……鼻音

破……無声破裂音

摩……無声摩擦音

“鼻—鼻”とは母音が前後を鼻音にはさまれた位置にあることを示す。

“単独”とは母音の位置がa.のいずれか1つの条件のみ満足していることを示す。

残りの84.3%はVCV音節標準パターンとのマッチングが正しくおこなえなかったことによる誤りである。認識誤りの原因となったVCV音節中の子音のconfusion matrixを表5.11に示す。この誤りの様子は第3章で述べたVCV音節の認識誤りの様子とよく似ている。これは認識誤りには特定の子音の誤りが影響しているのではなく，いずれの子音も同じように影響していることを意味している。したがってこの誤りを少なくするには，VCV音節全体の認識率の向上をはかることが必要である。

表5.11　認識誤りの原因となったVCV音節のconfusion matrix

	m	n	b	d	g	r	z	p	t	k	s	h	•
m		2	3		1	1							2
n	2		2			1	1						
b					3		1						1
d													
g	1	1	1	2									
r			5				2						1
z						1					1		
p													
t			2	2									
k							1		4				
s									1	1			
h		1		1	3				1		1		1

5.5.5　単語音声認識システムと本システムの音響処理結果の比較

前章で述べた単語認識システムと本連続単語システムの音響処理結果を比較する。セグメント認識部の構成は基本的には変っていないからセグメント化部の性能を比較する。

代表的な発声者TについてVCV音節の抽出結果を表5.12に示す。表5.12において“unique”とはVCV音節の抽出が一意におこなえたもの，“not unique”とは抽出が一意におこなえなかったものである。抽出が一意でないものは24.0%あるが，この場合2通りの抽出法があるものがほとんどである。したがってこのことは全体の処理量が抽出のあいまいさのため24%増すこ

表 5.12　VCV音節の抽出結果（発声者 T）

対象	correct		error	計
	unique	not unique		
2 連続単語	219 (77.1%)	51 (17.9%)	14 (4.9%)	284 (100%)
	270 (95.1%)			
3 連続単語	306 (68.3%)	118 (26.3%)	24 (5.4%)	448 (100%)
	424 (94.6%)			
4 連続単語	437 (66.7%)	164 (25.0%)	54 (8.2%)	655 (100%)
	601 (91.8%)			
計	962 (69.4%)	333 (24.0%)	92 (6.6%)	1,387 (100%)
	1,295 (93.4%)			

とを意味する。またセグメント化の誤りは92個(6.6％)あるが，表 5.8 を見るとわかるように実際認識誤りを引きおこすのは19個であり，VCV 音節全体の1.4％にすぎない。これに対し単語認識システムでは誤りは0.6%(21／3,850) と非常に少ないがあいまいさが極めて多い。すなわち単語 1 個当り 2.7 個の候補が生じており，このことは処理量が170％増すことを意味している。以上の結果をまとめて表 5.13 に示す。すなわち本システムのセグメント化部は誤りを少し許すかわりにあいまいさを大幅に減らすことに成功しており，性能はかなり改善されたといえる。

表 5.13　単語認識システムと本システムの音響処理結果の比較

比較事項	あいまいさを許したための処理量の増加率	セグメント化の誤り
単語認識システム	170 %	0.6 %
本システム	24 %	6.6 % 1.4％ (認識結果に影響するもの)

5.5.6　無声化の対策の効果

5.5.2 で述べた無声化の対策がどの程度効果的かを調べるため，無声化の生じた単語数とそのうち正しく認識された単語数を調べて表 5.14に示す。無声化は全体で59個生じているが，そのうち 81.4％は救うことができた。しかも表 5.3 に示した内容を辞書につけ加えることによる

認識誤り等の副作用は全く生じていない。あらかじめ無声化の規則を知り，その対策を施しておくことは言語情報の利用と考えられる。すなわち，ここで得られた結果は言語情報のような上位レベルの知識の利用が認識に極めて有効であることを示している

表 5.14　無声化の対策の効表

発　声　者	正しく認識されたもの / 無声化の生じた回数	%
T	15/19	78.9 %
I	15/19	78.9 %
S	18/21	85.7 %
計	48/59	81.4 %

5.5.7　まとめ

以上述べてきた検討事項をまとめて表 5.15に示す。

5.6　あとがき

以上述べてきたように，本章では，VCV音節を単位として連続単語音声を認識するシステムを構成し，90種類の日本語単語を 2 ～ 4 個連続して発声した90種類の連続単語を認識対象として認識実験を行った。その結果，3 名の発声者に対する 540サンプルを用いて 92.2 % の単語認識率が得られた。

本章で得られた成果をまとめると以下のようになる。

(1)　VCV音節を単位とした音響処理方法の精密化をはかった。単語音声認識の際用いた音響処理方法に，スペクトル変化，母音系列等の情報の利用を追加し，性能を向上させた。これにより，会話音声認識を行う際の音響処理方法に対する見通しが得られた。

(2)　新しい連続単語音声認識方法を提案した。従来の方法は，音声の先頭から 1 語ずつ認識

表5.16 検 討 事 項 の ま と め

項　　目	結　　論	根　　拠	対　　策
○単語の位置と認識率の関係	○語頭，語尾の単語の認識率は同じ程度 ○語中の単語はそれらに比較して認識率が低下する。	○実験結果および理論的検討	
○単語数と認識率との関係	○連続した単語数が増えても認識率の低下はわずか ○認識率$=\dfrac{2P_1+(n-2)P_2}{n}$ （P_1：語頭，語尾の単語の認識率） （P_2：語中の単語の認識率）	○実験結果 ○理論式と実験結果の一致	
○誤りのタイプ別分類	○分裂，置換の誤りはほぼ同じ程度 ○合同の誤りは非常に少ない	○実験結果および連続単語認識モデルによる定性的説明	
○認識誤りの原因の分析 a. セグメント化の誤り	○子音区間を母音区間と判定した音韻区間の分類誤りが最も多い（セグメント化の誤りの45.5％）	○実験結果	
	○これらの誤りは鼻音，摩擦音に特に多い	○子音区間が長い定常部が存在する	○分類の際スペクトル情報も用いる

項　　目	結　　果	根　　拠	対　　策
b．セグメントの認識誤り	○母音の認識誤りがかなり存在する（セグメントの認識誤りの25.3％）	○実験結果	
	○母音の認識誤りは鼻音の前後，無声摩擦音の前後，無声破裂音の後で生じやすい	○調音結合による母音のくずれ	○認識誤りのおこりやすい所では子音認識のあとで再び母音認識をおこなう
	○残りの74.7％はVCV音節標準パターンとのマッチングの誤りである	○実験結果	○VCV音節の認識論理の再検討
	○これらの誤りに特定の傾向はみられない	○実験結果	
○単語認識システムとの音響処理性能の比較	○性能はかなり改善された	○あいまいさによる処理量の増加率 170％から24％への減少 ○セグメント化の誤り 0.6％から1.4％へ増加	
○無声化の対策の効果	○辞書の内容の変更による方法が非常に有効である	○無声化の誤りの81.4％が救えた	

してゆく方法であり，1度誤りが生じるとその回復が困難であったのに対し，本章で提案した方法は，音声区間全体とすべての単語系列とのマッチングを行い，その中で最大の類似度が得られた単語系列を認識結果とする方法である。しかも，DPを用いることによって処理量の増加をおさえることを示した。この方法は，高い認識性能が得られるため，その後，各研究機関が採用しており，[71][120] 現在では連続単語音声認識の基本的な方法として定着した観がある。さらに，この方法をもとに，処理量の削減等も試みられている。[121]

認識性能をさらに向上させるには，VCV音節の認識率を高める必要がある。この問題は次章で取り上げる。

第6章　ＶＣＶ音節の音声認識法の改良

6.1　はしがき

第３章～第５章では，VCV音節の認識，VCV音節を単位とした単語音声の認識，VCV音節を単位とした連続単語音声の認識について述べた。これらの結果により，VCV音節を単位にとることの有効性が示されたが，次のような検討事項が残されていることも明らかになった。すなわち，単語音声認識，連続単語音声認識における誤りを調べると，VCV音節の認識誤りに起因したものがかなりの割合を占める。したがって，単語音声認識，連続単語音声認識の性能の向上をはかるためには，VCV音節の認識法の改善が必要である。そこで，本章では，VCV音節の認識法の改良を取り上げ検討を行う。まず，VCV音節の認識系について述べた後，VCV音節標準パターンの作成法について述べ，２種類の標準パターンを提案する。次に，DPマッチング法によるVCV音節の認識法，および，いくつかの改良案について述べる。さらに，それらの標準パターンと認識法を組み合わせて行ったいくつかの認識実験について述べると共に，認識実験結果について検討を行う。最後に，改良した認識法を連続単語音声に適用した実験，およびその結果について述べる。

6.2　認識系の構成

6.2.1　認識対象

第３章では，VCV音節の認識を行い，高い認識率が得られることを示した。ただし，この実験において，認識対象は単独に発声したVCV音節であった。しかしながら，最終目標が会話音声の認識である以上，この認識実験だけでは不十分であり，連続音声中のVCV音節を対象とした認識実験を行う必要がある。そこで，本章では，種々の長さの連続音声を対象とすることにする。そのうちわけを次に示す。

単語
- VCV型単語 ……30種類 （6.1）
- VCVCV型単語 ……30種類 （6.2）
- VCVCVCV型単語……30種類 （6.3）

連続単語
- 単語を2個連続して発声したもの……30種類 （6.4）
- 単語を3個連続して発声したもの……30種類 （6.5）
- 単語を4個連続して発声したもの……30種類 （6.6）

（6.1）～（6.3）の単語は，日本語の有意味単語から適当に選んだものであり，第4章で述べた単語音声認識の際，認識対象としたものと同じである。（6.4）～（6.6）の連続単語は，（6.1）～（6.3）の単語からランダムに選んだ単語をつなぎあわせたものであり，第5章で述べた連続単語音声認識において用いたものと同じである。

これらに含まれる音韻は，次に示す母音5種類，子音12種類である。

母音…… / a, i, u, e, o /

子音…… / m, n, b, d, g, r, z, p, t, k, s, h /

6.2.2 認識系

入力音声の特徴抽出は第2章で述べたのと同じものを用いる。入力音声を表現するパラメータとしては，15 msecのフレームごとに得られる音声波形の p 次までの自己相関関数

$$\boldsymbol{v} = (v_0, v_1, \cdots\cdots, v_p) \qquad (6.7)$$

を用いる。入力のVCV音節は，vの時系列

$$\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\cdots, \boldsymbol{v}_N) \qquad (6.8)$$

$$\boldsymbol{v}_i = (v_{i0}, v_{i1}, \cdots\cdots, v_{ip}) \qquad (6.9)$$

として表わされる。ただし，Nは入力のフレーム数であり，$\boldsymbol{v}_i$は第 i フレームの自己相関関数である。これに対し，VCV音節標準パターンは，最尤スペクトルパラメータ

$$\boldsymbol{A} = (A_0, A_1, \cdots\cdots, A_p) \qquad (6.10)$$

の時系列

$$(A_1, A_2, \cdots\cdots, A_M) \tag{6.11}$$

$$A_j=(A_{j0}, A_{j1}, \cdots\cdots, A_{jp}) \tag{6.12}$$

として表現される。ただし，Mは標準パターンのフレーム数であり，A_jは第jフレームの最尤スペクトルパラメータである。また，VCV音節標準パターンの種類は300個（母音5×子音12×母音5）である。

VCV音節の認識を行うには，まず，入力音声とVCV音節標準パターンとのパターンマッチングを行い，それらの間の類似度を計算する。次に，類似度が最大であるVCV音節のカテゴリーに入力が属するものと決定する。標準パターンの作成法，認識方法については6.3節，6.4節で述べる。

6.3 標準パターン作成法[123]

6.3.1 学習サンプルの非線形な平均

一般に，音声は発声するたびにそのパターンが変動する。したがって，未知サンプルに対して適応性のある標準パターンを作るには，同じカテゴリーに属するいくつかの学習サンプルを平均する方法が有効である。平均する際，各学習サンプルが時間軸方向に非線形な伸縮をしていることを考慮する必要がある。そのため，時間軸の非線形な伸縮を補正した後，学習サンプルを平均するという方法をとる。具体的な作成法は第3章でくわしく述べてあるので，ここでは省略する。本実験においては，学習サンプルの選び方によって，以下に述べる2種類の標準パターンを用いる。

6.3.2 標準パターンA

学習サンプルとして，単独に発声したVCV音節を用いて標準パターンを作成する。この方法は，学習サンプルの収集，標準パターンの作成が比較的容易であるという利点がある。

6.3.3 標準パターンB

学習サンプルとして，連続音声中のVCV音節を用いて標準パターンを作成する。連続音声中のVCV音節は，調音結合の影響を受け，単独のVCV音節に比べ変形していると考えられる。したがって，連続音声中のVCV音節を認識する際は，単独発声のVCV音節から作成した標準パターンより，連続音声中のVCV音節から作成した標準パターンの方が好ましいと考えられる。ここでの標準パターン作成は，このような考え方に立っている。

連続音声中のVCV音節は，前後の音韻の影響を受けているため，その影響を取り除くには，種々の音韻構成の学習サンプルから標準パターンを作成し，前後の音韻の影響を平均化する必要がある。具体的な作成法は次の通りである。

(1) 学習サンプル作成のための音声サンプルは，VCVCVCV型の無意味単語とする。この無意味単語は，VCV音節を3個つないだものであるから，連続音声中のVCV音節という条件を満足した学習サンプルを抽出することができる。かつ，あまり長い無意味単語は発声しにくいため，適当な長さということで，この型を選んだ。

(2) 標準パターンは300種類必要である。そのため，一組の学習サンプルを作成するには，すべてのVCV音節が各1回，しかもランダムに出現するような100個のVCVCVCV型単語を用いれば良い。

(3) 上の条件を満足するVCVCVCV型単語の母音部分を決定するために，まず，/aa/，/ai/，/au/，/ae/，/ao/，/ia/，……，/oo/の25種類の母音対にそれぞれ数字の1～25を割り当てる。これらの数字を3個組み合わせると，VCVCVCV型単語の母音部分が決定する。ただし，組み合わせる数字には制限が課せられる。(たとえば，母音対/ai/に接続できる母音対は，/ia/，/ii/，/iu/，/ie/，/io/のみである。) これを定式化すると次のようになる。

数字nの次には

$$\mathrm{mod}\,(n-1,\ 5)\times 5 < m \leqq \mathrm{mod}\,(n,\ 5)\times 5 \qquad (6.13)$$

($\mathrm{mod}\,(i,\ j)$はiをjで割った余りを示す)

を満足する数字mのみが接続できる。

上の条件を満足する3つ組の数字を100個作り，これらの中には25種類の数字が各12回づつランダムに出現するようにする。こうして作った3つ組の数字の例を表6.1に示す。これから

100個のVVVV型単語が出来る。

表 6.1 標準パターンB作成用の数字の3つ組の例

12. 6. 2	8. 11. 5	23. 15. 22	9. 18. 14	17. 7. 10
21. 1. 4	19. 20. 25	24. 16. 3	13. 15. 22	8. 11. 5
21. 3. 14	19. 17. 9	20. 25. 24	16. 2. 6	1. 4. 18
13. 12. 7	10. 23. 13	14. 18. 15	25. 22. 6	2. 9. 16
1. 5. 23	12. 7. 10	24. 17. 8	11. 4. 19	20. 21. 3
15. 25. 24	19. 20. 23	12. 10. 22	9. 18. 11	1. 2. 8
13. 14. 17	7. 6. 4	16. 5. 21	3. 13. 12	7. 6. 4
19. 20. 24	18. 15. 21	5. 22. 9	17. 10. 25	23. 11. 1
3. 14. 16	2. 8. 13	12. 6. 4	16. 3. 15	23. 11. 5
22. 10. 25	24. 17. 7	9. 19. 18	14. 20. 21	1. 2. 8
11. 4. 19	20. 22. 9	18. 14. 16	2. 7. 6	1. 5. 24
17. 8. 13	15. 23. 12	10. 25. 21	3. 14. 19	18. 15. 24
16. 4. 17	6. 2. 7	9. 20. 23	13. 12. 10	21. 1. 3
11. 5. 25	22. 8. 15	23. 11. 5	25. 21. 4	17. 9. 16
1. 3. 13	12. 10. 24	18. 14. 19	20. 22. 7	6. 2. 8
15. 25. 21	5. 22. 6	2. 7. 10	23. 11. 4	17. 9. 19
20. 24. 18	12. 8. 13	14. 16. 1	3. 14. 17	7. 9. 19
16. 2. 10	24. 18. 11	4. 20. 25	21. 1. 5	23. 12. 8
13. 15. 22	6. 3. 11	3. 15. 22	9. 17. 6	2. 8. 14
20. 21. 1	4. 16. 5	25. 24. 19	18. 13. 12	7. 10. 23

(4) 上で作成したVVVV型単語のVとVの間にCを挿入する。このとき，25種類のVVの組のそれぞれに対して，12種類の子音を各1回ランダムに挿入する。このようにして100個のVCVCVCV型単語が作成される。その例を表 6.2 に示す。

(5) 以上の手順で作成された100個のVCVCVCV型単語中には，300種類のVCV音節が各1個存在し，しかもその位置はランダムである。このような100個のVCVCVCV型単語の組を複数個用意し，各音声サンプルから視察によって切り出したVCV音節を学習サンプルとして，標

表 6.2　標準パターンB作成用のVCVCVCV型単語の組の例

UZITADI	INUDATO	OTUTOGI	IGEGUTE	ENINIZO
OZAZAGE	EZEDONO	OTENAZU	UGUGODI	IKUHANO
ORARUNE	EMEKIKE	EHOKOGE	EKAHIGA	AMADEKU
UDURIKI	IMONUKU	UPERUHO	OHOKIHA	AZIREHA
AKAPOPU	UKIHIKO	OHEMIMU	UZAKEKE	EZOKAKU
UBOGOBE	EBESOBU	UDIBOZI	IBEBUSA	ABASIZU
UZUBEZI	IGIBAZE	EGABODA	ADUBUSI	ISIMABE
EPETOME	EMUMOSA	AHOBIME	EPIPOSO	OHUTAPA
ASUHESA	ATIPUNU	UGIZANE	ERAGUZO	OZUNAZO
ONIGORO	OZEBIRI	IDEGEDU	UZENOGA	AGANIBU
URAHENE	EROHINE	ENUKEBA	ARIBIPA	ANAKOPE
ERIRUHU	UPOKUPI	INOBOPA	APUDETE	EHUNONE
EPAMESI	INAMIPI	IHEMODU	UMUTITO	OTATATU
UMADOPO	OMISURO	OGUBAGO	OMOMAPE	ETIZEMA
AHAMUPU	UMIHORE	EZUGEHE	EBOPIMI	IRABITU
UKOTOHA	AMOSIKA	APITIRO	OMUPASE	EHITERE
EPOKETU	UHIHUSU	UMETARA	AHUSEDI	IDIPEDE
EDAKIDO	OSEPUKA	ATEKODO	OBADASO	OSUBIDU
UTUSOTI	ISABUGA	ANUDORI	ISEGIDA	AGIGURE
EGONASA	AREZARO	OZODESE	ESURUNI	IZISORU

準パターンを作成する。

以上の手順で，前後の音韻の影響を平均化した標準パターンを作成することができる。しかしながら，このような標準パターンは，標準パターンAに比較して，実際面で次のような欠点を持つ。

(1)　発声者にとって無意味単語は発声しにくいため，音声サンプルの収集に手間がかかる。

(2)　大量の音声サンプルから視察によりVCV音節を切り出すため，標準パターンの作成に手間がかかる。

6.4 認識法[123]

6.4.1 DPマッチング法

6.3.1でも述べたように，音声は発声するたびに時間軸の非線形な伸縮が生じる。したがって，入力音声と標準パターンのマッチングを行う際には，時間軸の正規化を行う必要がある。マッチングの具体的な方法は次の通りである。

入力音声を

$$\boldsymbol{V}=(\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots\cdots,\ \boldsymbol{v}_N) \tag{6.14}$$

$$\boldsymbol{v}_i=(v_{i0},\ v_{i1},\ \cdots\cdots,\ v_{ip}) \tag{6.15}$$

VCV音節標準パターンを

$$(\boldsymbol{A}_1,\ \boldsymbol{A}_2,\ \cdots\cdots,\ \boldsymbol{A}_M) \tag{6.16}$$

$$\boldsymbol{A}j=(A_{j0},\ A_{j1},\ \cdots\cdots,\ A_{jp}) \tag{6.17}$$

とする。式（6.14），（6.15）より，類似度マトリクスLMを作成する。

$$LM=\{\ l(i,\ j)\ \} \tag{6.18}$$

式（6.18）において，$l(i,\ j)$は入力の第iフレームと標準パターンの第jフレームの類似度であり，次式で表現される。

$$l(i,\ j)=-\log\left\{\sum_{k=0}^{p}A_{jk}\,v_{ik}\right\} \tag{6.19}$$

時間軸の非線形な伸縮を考慮したマッチングを行うために，次式の類似度和を求める。

$$\max_{f}\left\{\sum_{i=1}^{N}l(i,\ f(i))\right\} \tag{6.20}$$

ここで，fは時間軸の変換を行う関数であり，音声の時間軸伸縮の仕方にかなった適当な制約条件をつける必要がある。第3章では次の条件を用いた。

(i) $f(1)=1,\qquad f(N)=M$ (6.21)

(ii) $$f(i)=\begin{cases}f(i-1)\\ f(i-1)+1\\ f(i-1)+2\end{cases} \tag{6.22}$$

（6.21）は，入力と標準パターンの始端，終端を一致させる条件であり，これを「端点条件」と呼ぶことにする。また，式（6.22）は時間軸の伸縮に制限をつける条件で，これを「自由度」と呼ぶことにする。端点条件．自由度共に，ここで示したものが唯一というわけではなく，後で述べるように，いくつかの変形が可能である。

式（6.20）は，類似度マトリクス上で，要素（1，1）から（N，M）へ至る類似度和が最大のパスを探索する問題になっており，ダイナミック・プログラミング（DP）を用いて，次のように能率良く求めることができる。

まず，$L(i, j)$ を次のように定義する。

$$L(i, j) = \max_{f} \left\{ \sum_{k=1}^{i} l(k, f(k)) \right\} \qquad (6.23)$$

ただし，f は次の条件を満す。

$$\left.\begin{aligned} &f(1)=1 \qquad f(i)=j \\ &f(k)=\begin{cases} f(k-1) \\ f(k-1)+1 \\ f(k-1)+2 \end{cases} \end{aligned}\right\} \qquad (6.24)$$

（6.23）は，入力が（$\boldsymbol{v}_1$，$\boldsymbol{v}_2$，……，$\boldsymbol{v}_i$），標準パターンが（$\boldsymbol{A}_1$，$\boldsymbol{A}_2$，……，$\boldsymbol{A}_j$）のときの両者の類似度であるから，元の入力，および標準パターン（式（6.14），（6.16））の部分類似度である。$L(i, j)$ は，漸化式

$$L(i, j) = \max \begin{bmatrix} L(i-1, j) + l(i, j) \\ L(i-1, j-1) + l(i, j) \\ L(i-1, j-2) + l(i, j) \end{bmatrix} \qquad (6.25)$$

を順次とくことによって求められる。そして，（6.20）は次式で与えられる。

$$L(N, M) = \max_{f} \left\{ \sum_{i=1}^{N} l(i, f(i)) \right\} \qquad (6.26)$$

また，時間軸の変換関数 f は，$L(N, M)$ を計算する際加えた $l(i, j)$ の位置を逆にたどることによって求めることができる。図 6.1 にこのマッチングの様子を示す。このようにして求められた f を f^* とする。このとき，

$$\sum_{i=1}^{N} \{ l(i, f^{*}(i)) \} \qquad (6.27)$$

が入力と標準パターンとの間の類似度和であり，認識結果は，最大の類似度和を与える標準パターンが属するカテゴリーとする。ただし，端点条件，自由度と同様，式（6.27）についても後で述べるような変形が可能である。

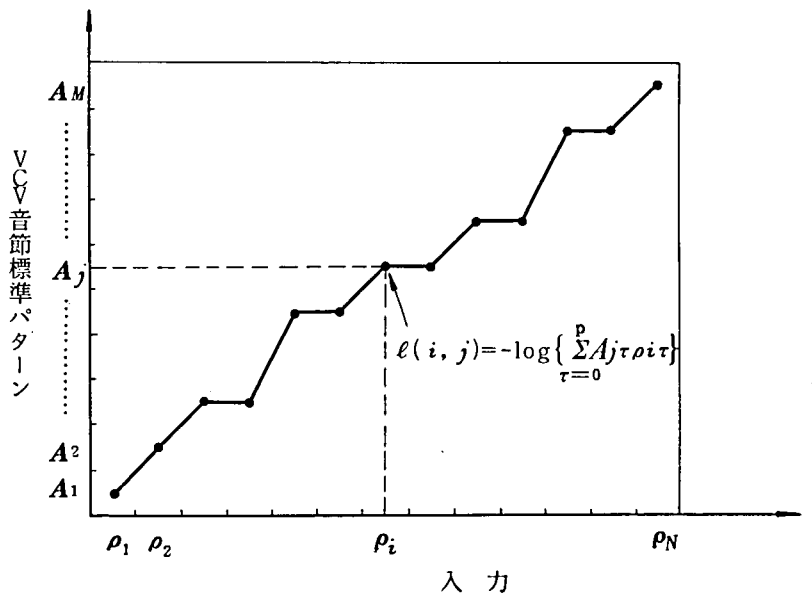


図 6.1　DPマッチングによるVCV音節の認識

6.4.2　認識法の改良

6.4.1で述べたDPマッチング法に対し，認識率の向上のためいくつかの変形をおこなう。

6.4.2.1　端点条件

先に述べたように，式（6.21）の条件は，入力と標準パターンの始端，終端を一致させる条件であった。しかしながら，単独に発声したVCV音節から作成した標準パターンと連続音声中のVCV音節とのマッチングの様に，両方のパターンの性質が異なっている場合には，もっとゆるやかな条件の方が良いと考えられる。そこで，式（6.2）の変形として次の条件を提案する。

$$1 \leq f(1) \leq f(N) \leq M \qquad (6.28)$$

この条件を用いると，入力と標準パターンの端点は必ずしも一致しない。以後は式（6.21）の条件を「端点固定」，式（6.28）の条件を「端点フリー」と呼んで区別する。図6.2に端点フリーの条件におけるDPパスの様子を，また，図6.3に入力と標準パターンのマッチングの様子を示す。図6.3からわかるように，端点フリーの方法では，入力が標準パターンの一部に対応づけられる。

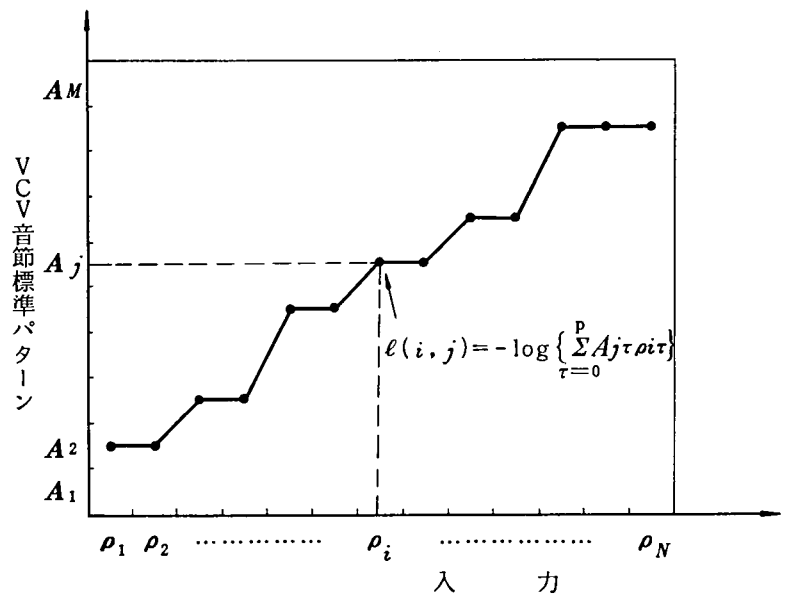


図6.2　端点フリーDPマッチングのパス

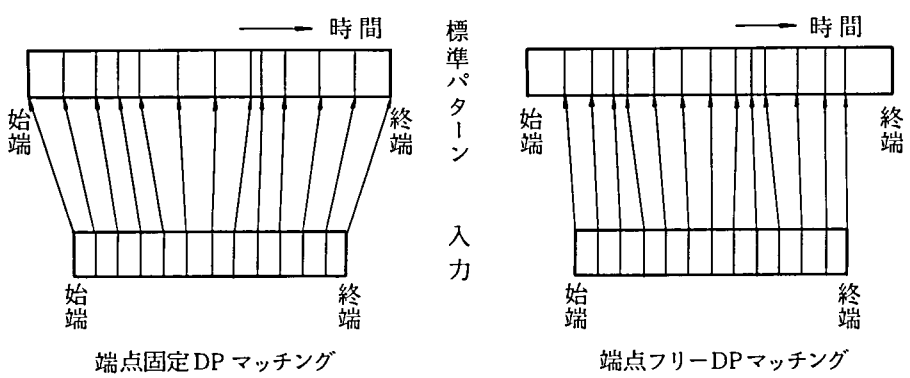


図6.3　入力と標準パターンの対応づけの比較

端点固定DPマッチング……入力と標準パターンの始端同士，終端同士を対応づける。
端点フリーDPマッチング……入力の始端，終端は必ずしも標準パターンの始端，終端に一致させない。

6.4.2.2 パスの自由度

条件式（6.22）は，DPパス上の各点において，進む方向に3つの自由度がある。自由度を大きくすると，時間軸の正規化の能力を増すことになるが，同時に，異なった音韻間のconfusionを増大させる原因ともなる。したがって，ここでは，パスの自由度を式（6.22）よりきびしくした条件と，ゆるやかにした条件を提案する。

まず，式（6.22）よりきびしい条件として，

$$\left.\begin{array}{ll} f(i)=1 \text{ の時} ; & f(i+1)=1, 2 \\ f(i)=M \text{ の時} ; & f(i+1)=M \\ \text{その他の時} ; & f(i+1)=f(i)+1 \end{array}\right\} \qquad (6.29)$$

を用いる。これは，$f(i)=1$ の場合を除いて自由度が1であるため「自由度1の条件」と呼ぶ。端点フリーの条件と組み合わせて用いた場合の入力と標準パターンのマッチングの様子を図6.4に示す。これは，入力と標準パターンの時間軸をずらしながら線形マッチングを行い，類似度

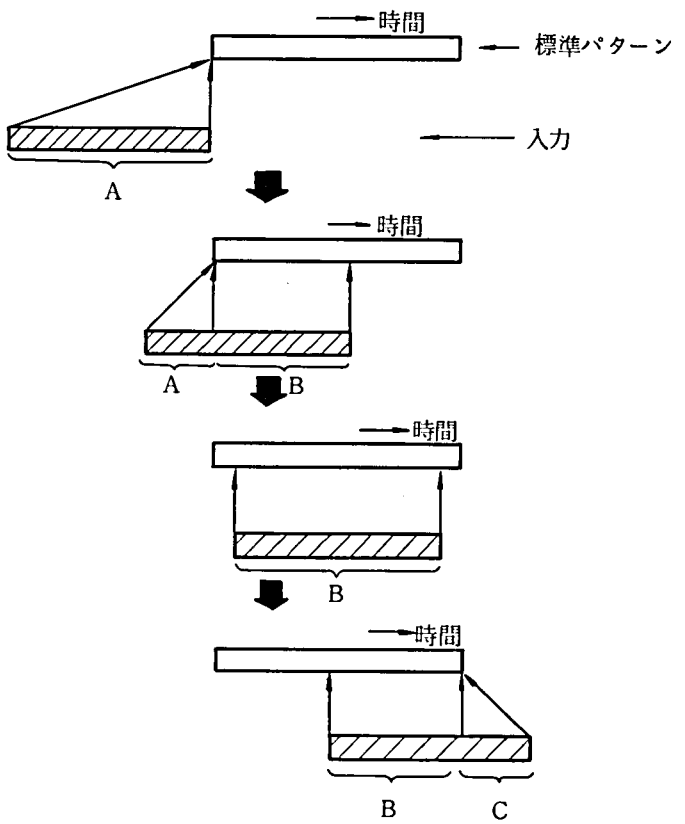


図6.4　線形マッチングにおける入力と標準パターンの対応づけ

A ……標準パターンの第1フレームに対応づけられる。
B ……標準パターンの各フレームと1対1に対応づけられる。
C ……標準パターンの最終フレームに対応づけられる。

和の最大値を求めることに相当する。この場合には，実質的には時間軸の非線形正規化を行っていないため，「線形マッチング」と呼ぶことにする。

次に，パスの自由度が式（6.22）よりゆるやかな条件として，

$$f(i)=\begin{cases} f\ (i-1) \\ f\ (i-1)+1 \\ f\ (i-1)+2 \\ f\ (i-1)+3 \end{cases} \qquad (6.30)$$

を用いる。この場合，DPパス上の各点においてパスの自由度は4であるため「自由度4の条件」と呼ぶことにする。これらの条件に対し，式（6.22）は「自由度3の条件」と呼ぶことにする。

以上の3種類のパスの自由度を比較して図6.5に示す。

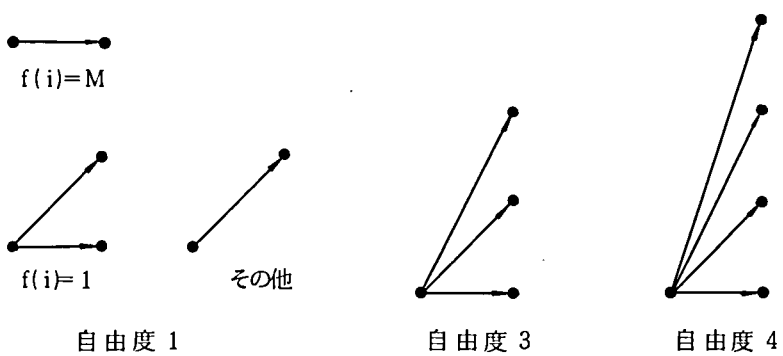


図6.5　DPにおけるパスの比較

6.4.2.3　類似度和を計算する際の重みづけ

式（6.27）においては，類似度和を求める際，各類似度をそのまま加えている。これに対し，重みづけをして加えることが考えられる。種々の重みづけが考えられるが，ここでは次の方法を提案する。

$$\sum_{i=k+1}^{N-k} \left\{ l\ (i,\ f^{*}(i)\ ) \right\} \qquad (6.31)$$

これは，入力の語頭，語尾のkフレームの類似度の重みを0としたものであり，いいかえれば，語頭，語尾のkフレームを無視することに相当する。

以上述べた，標準パターン，および，認識法の種類をまとめて表 6.3 に示す。

表 6.3　標準パターン，認識法の種類

		種　　類
標準パターン		A　　B (単独のVCV音節から作成)，(連続音声中のVCV音節から作成)
認識法	端点条件	固定，フリー
	自由度	1，3，4
	類似度和の際の重みづけ	無，有

6.5　認識実験[123]

6.5.1　実験に用いた音声サンプル

5 名の発声者（T，S，I，H，K）による次の 6 組の音声サンプルを認識実験に用いた。

サンプル a …… VCV型単語30種類を各 5 回発声したもの。

b …… VCVCV型単語30種類を各 2 回発声したもの。

c …… VCVCVCV型単語30種類を各 2 回発声したもの。

d …… 2 連続単語30種類を各 1 回発声したもの。

e …… 3 連続単語30種類を各 1 回発声したもの。

f …… 4 連続単語30種類を各 1 回発声したもの。

各サンプルからのVCV音節の切り出しは視察により行った。各発声者ごとのVCV音節のサンプル数は次の通りである。

サンプル a …… 150 個

サンプル b …… 120 個

サンプル c …… 180 個

サンプルd …… 120個

サンプルe …… 180個

サンプルf …… 246個

6.5.2 実験の説明

標準パターン，認識法を種々組み合わせて，計6種類の実験を行った。これらを，実験1，実験2，……，実験6と呼ぶことにする。各実験の比較が行いやすいように，表6.4に実験の分類を示してある。ただし，これらの実験において，入力のVCV音節の母音の種類については

表6.4 実 験 の 分 類

実験番号	標準パターン	認識法			発声者	音声サンプル	表
		端点条件	自由度	類似度和の際の重みづけ			
1	A	固 定	3	無	T,S,I,H,K	a～f	6.5
2	〃	〃	〃	有	〃	a	6.6
3	〃	フリー (線形マッチング)	1	〃	〃	a	6.7
4	〃	固 定 フリー	3 4	無	〃	d	6.8
5	〃	フリー	3	有	〃	d～f	6.9
6	B	固 定	3	無	T	a～f	6.10

既知であるとした。したがって，認識の際は子音部の異なる12種類のVCV音節標準パターンとマッチングを行うことになる。各実験の結果は表6.5～表6.10に実験ごとにまとめてある。

表6.5　実　験　1　の　結　果

（端点固定DPマッチングによる認識実験）

発声者＼サンプル	a	b	c	d	e	f
T	93.3％	84.2％	70.0％	73.3％	75.0％	77.2％
S	81.3％	80.0％	73.9％	73.3％	70.6％	64.2％
I	72.0％	69.2％	67.2％	75.0％	63.8％	61.5％
H	84.0％	64.1％	48.3％	53.3％	41.7％	44.7％
K	70.7％	73.3％	60.0％	43.3％	47.8％	45.1％
平　均	80.3％	74.2％	63.9％	63.7％	59.8％	58.6％

表6.6　実　験　2　の　結　果

（端点固定，類似度和の重みづけを行った認識実験）

発声者＼k*	0	1	2	3	4	5	6	7	8	9	10	11
T	93.3％	93.3％	94.7％	94.0％	93.3％	93.3％	93.3％	92.7％	90.0％	84.7％	78.0％	54.7％
S	81.3％	82.7％	84.0％	84.7％	86.0％	86.0％	88.0％	88.7％	82.7％	76.0％	62.7％	46.7％
I	72.0％	72.7％	75.3％	78.0％	80.0％	80.7％	80.7％	80.0％	80.7％	78.0％	75.3％	68.7％
H	84.0％	84.0％	84.7％	84.7％	83.3％	84.7％	81.3％	76.7％	70.7％	56.0％	36.7％	12.7％
K	70.7％	74.0％	74.7％	76.7％	76.0％	78.7％	78.7％	80.0％	78.0％	77.3％	82.0％	73.3％
平　　均	80.3％	81.3％	82.7％	83.6％	83.7％	84.7％	84.4％	83.6％	80.4％	74.4％	66.9％	51.2％

（*k は，類似度和の計算の際無視する語頭，語尾のフレームの数）

表6.7　実　験　3　の　結　果

（線形マッチングによる認識実験）

発声者 \ k*	0	1	2	3	4	5	6	7	8	9	10	11
T	88.7％	90.0％	92.0％	92.0％	92.7％	90.7％	91.3％	88.7％	83.3％	78.0％	69.3％	46.7％
S	88.0％	90.0％	90.7％	90.0％	90.7％	88.7％	86.0％	86.7％	81.3％	75.3％	58.0％	41.3％
I	72.7％	73.3％	74.7％	75.3％	74.7％	74.0％	74.7％	74.0％	70.7％	72.7％	75.3％	70.7％
H	76.0％	74.7％	74.7％	72.7％	72.7％	73.3％	73.3％	66.7％	62.7％	46.7％	33.3％	15.3％
K	61.3％	64.0％	65.3％	62.0％	66.0％	68.7％	71.3％	69.3％	68.7％	68.0％	65.3％	66.0％
平　　均	77.3％	78.4％	79.5％	78.4％	79.3％	79.1％	79.3％	77.1％	73.3％	68.1％	59.9％	48.0％

（*kは，類似度和の計算の際無視する語頭，語尾のフレームの数）

表6.8　実　験　4　の　結　果
（端点条件，自由度の比較実験）

端点 / 自由度 / 発声者	固　定	固　定	フ　リ　ー	フ　リ　ー
	3	4	3	4
T	73.3％	76.7％	79.2％	72.5％
S	73.3％	69.2％	71.7％	70.8％
I	75.0％	79.2％	82.5％	81.7％
H	53.3％	53.3％	55.0％	58.3％
K	43.3％	46.7％	46.7％	45.0％
平　　均	63.7％	65.0％	67.0％	65.7％

表6.9　実　験　5　の　結　果

（端点フリー，類似度和の重みづけを行った認識実験）

発声者＼サンプル ＼ k*		0	1	2	3	4
T	d	79.2％	82.5％	81.7％	78.3％	74.2％
	e	78.9％	80.0％	79.4％	68.9％	64.4％
	f	80.1％	79.3％	79.7％	76.4％	67.5％
	全　体	79.5％	80.2％	80.0％	74.4％	67.9％
S	d	71.7％	78.3％	74.2％	71.7％	73.3％
	e	73.3％	75.0％	76.7％	76.1％	71.7％
	f	73.2％	74.4％	74.8％	69.5％	65.9％
	全　体	72.9％	75.5％	75.3％	72.2％	69.4％
I	d	82.5％	84.2％	84.2％	82.5％	78.3％
	e	71.7％	73.3％	73.9％	73.3％	71.1％
	f	75.8％	77.0％	74.6％	72.1％	70.5％
	全　体	75.9％	77.4％	76.5％	74.8％	72.4％
H	d	55.0％	53.3％	53.3％	48.3％	27.5％
	e	45.0％	46.7％	45.0％	42.8％	35.0％
	f	52.5％	50.0％	49.6％	48.4％	37.3％
	全　体	50.6％	49.6％	48.9％	46.5％	34.4％
K	d	47.5％	46.7％	46.7％	44.2％	44.2％
	e	51.7％	53.3％	49.4％	48.9％	46.7％
	f	52.8％	49.6％	51.6％	51.2％	52.8％
	全　体	51.3％	50.2％	49.8％	48.9％	48.9％
平　均	d	67.2％	69.0％	68.0％	65.0％	59.5％
	e	64.1％	65.7％	64.9％	62.0％	57.8％
	f	66.8％	66.0％	66.0％	63.5％	58.8％
	全　体	66.0％	66.5％	66.1％	63.3％	58.6％

（*kは，類似度和の計算の際無視する語頭，語尾のフレームの数）

表 6.10 実 験 6 の 結 果

(連続音声中のVCV音節より作成した標準パターンを使用)

サンプル	a	b	c	d	e	f
認識率	76.0 %	75.3 %	78.4 %	75.8 %	68.3 %	80.5 %

(発声者 T)

6.6 検 討

6.5 で述べた実験結果に基づいて種々の検討を行う。

6.6.1 標準パターンの検討

標準パターンA，Bの優劣を確かめるために，実験1と6の結果を比較する。発声者Tに関して，標準パターンA，Bを用いて認識した結果を比較して図 6.6 に示す。ただし，図 6.6 において，横軸はVCV音節を切り出した音声サンプルの長さをVCV音節単位の音節数で示してあ

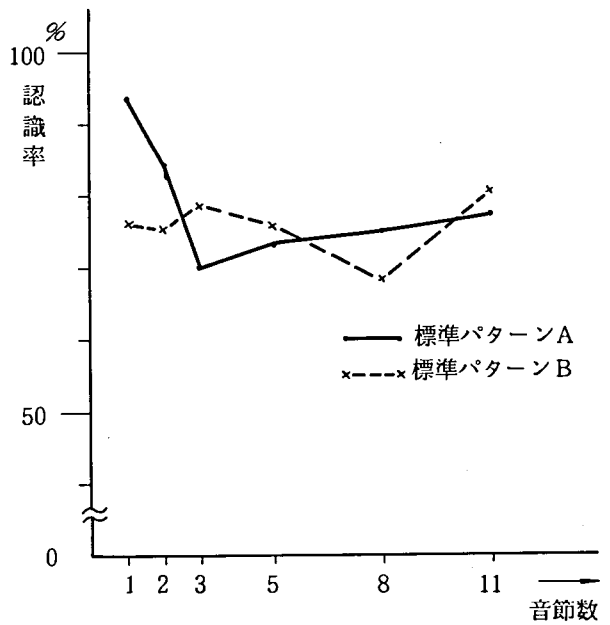


図 6.6 標準パターンA，Bによる認識結果の比較

る。たとえば，VCVCV型単語中から切り出したVCV音節の認識結果は，音節数２の位置に示してある。また，２～４連続単語は，それぞれ平均して５，８，11個のVCV音節から構成されているから，それらから切り出したVCV音節の認識結果は，それぞれ，音節数５，８，11の位置に示してある。図6.6の結果を見ると，標準パターンＢの場合は，どの音節数に対してもほぼ同じ程度の認識率である。これに対し，標準パターンＡを用いると，音節数１，２に対しては非常に良い認識率が得られているが，それ以外の音節数の場合は認識率が下がり，標準パターンＢと同じ程度の認識率になっている。音節数１，２の場合は，入力のVCV音節は前後の音韻による調音結合の影響をあまり受けていないので，標準パターンＡを用いると認識率が良いことは理解できる。これに対し，音節数３～11の場合に標準パターンＢが標準パターンＡと同じ程度の認識率しか得られないことは，連続音声中のVCV音節を認識する際，連続音声中のVCV音節から作成した標準パターンを用いても効果がないことを示している。したがって，標準パターン作成時の手間等も考えると，連続音声中のVCV音節を認識する際も単独のVCV音節から作成した標準パターン（標準パターンＡ）で十分であると結論できる。

6.6.2　認識法の検討

まず，最も基本的な方法として，端点固定，自由度３，重みづけなし，の場合の結果を検討する。実験１の結果を，発声者別，音節数別の認識率のグラフにして図6.7に示す。図6.7から次のことがわかる。

(1)　１VCV音節から成る音声サンプルの場合でも，発声者間で約20%の認識率の差がある。認識率の低い発声者は約70%の認識率しか得られておらず，まだ改善の余地がある。

(2)　入力音声の音節数が増加すると共に認識率が低下する。一般の連続音声はかなり多くの音節数からなるので，このような現象は好ましくなく，改善する必要がある。これらの問題点を，(1)１VCV音節からなる音声サンプルのように，母音定常部を含んだVCV音節の認識，(2)連続音声中から切り出されたVCV音節のように，母音定常部を含まないVCV音節の認識，に分けて，それぞれの場合について以下に検討を行う。

6.6.2.1　母音定常部を含むVCV音節の認識法の検討

(a)　DPマッチングの検討

母音定常部を含んだVCV音節の認識において，発声者間で認識率に大きな差が生じる原因の

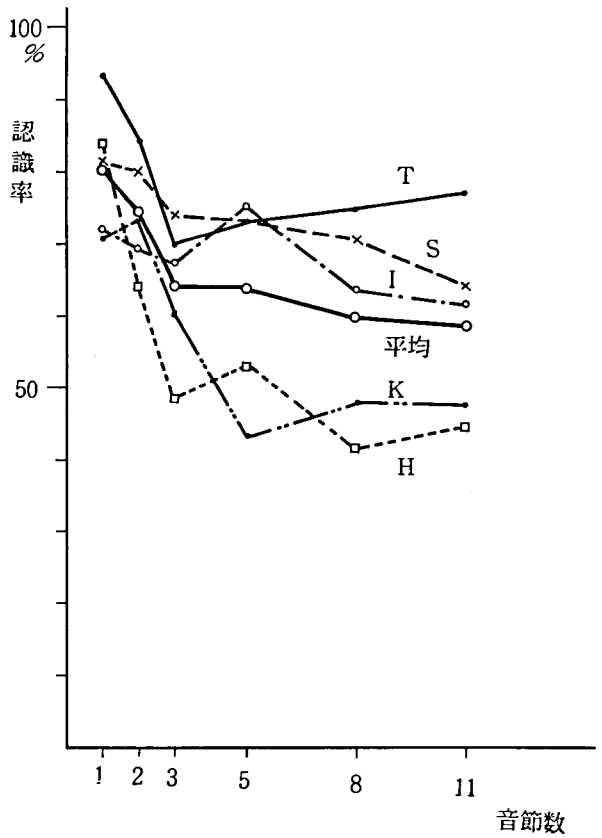


図 6.7　実験 1 の結果

1つとして，認識法（DPマッチング）に問題があるのではないかと考えられる。そこで，実験1の結果から，認識率の良い発声者（T）と悪い発声者（I）を選び，これら2名の発声者について，DPマッチングによる時間軸正規化の様子を検討する。

DPマッチングは，入力と標準パターンの間の時間軸を正規化する能力を持っているが，両者の時間構造の差が大きすぎると，正規化が十分行えずに認識誤りの原因となる可能性がある。そこで，発声者T，Iについて，入力と標準パターン間の時間構造の差を比較する。VCV音節の時間構造を示す指標として，入力と標準パターンの長さの比をとり，次式で定義する。

$$\left.\begin{aligned} R_1 &= \text{語頭の母音部における} \ \frac{\text{入力の長さ}}{\text{標準パターンの長さ}} \\ R_2 &= \text{子音部における} \ \frac{\text{入力の長さ}}{\text{標準パターンの長さ}} \end{aligned}\right\} \quad (6.32)$$

$$R_3 = \text{語尾の母音部における} \frac{\text{入力の長さ}}{\text{標準パターンの長さ}}$$

$$R_4 = \text{VCV音節全体における} \frac{\text{入力の長さ}}{\text{標準パターンの長さ}}$$

発声者T，Iについて，サンプルaの各30サンプルを用いて，R_1～R_4の平均値および標準偏差を求めた結果を表6.11に示す。また，時間構造の差と認識結果との相関を見るため，次式で定義される類似度差DとR_4の相関を求め，あわせて表6.11に示しておく。

表6.11　入力と標準パターンの長さの比

発声者	比	平均	標準偏差
I	R_1	1.12	0.18
	R_2	1.13	0.27
	R_3	1.45	0.42
	R_4	1.22	0.14
T	R_1	1.00	0.19
	R_2	0.99	0.32
	R_3	1.23	0.38
	R_4	1.05	0.12

発声者	I	T
R_4と類似度差の相関	−0.246	−0.260

D＝(入力と同じカテゴリーに属する標準パターンとの類似度和)−(入力と異なるカテゴリーに属する標準パターンとの類似度和の最大値)　(6.33)

表6.11を見ると，R_1～R_4の平均値が発声者Iの方が大きい。すなわち，発声者Iの方が，標準パターンに比べ入力が長いことを示している。しかしながら，R_4とDの相関は2人共ほとん

どないから，このような時間構造の差は認識結果と直接は関係ないといえる。すなわち，発声者T，Iにおいて，VCV音節の時間構造にある程度の差はあるが，それが認識誤りの原因とはいえないと結論できる。

次に，より直接的にDPマッチングの良否を判定するため，標準パターンと入力音声の対応づけの良否を調べる。標準パターンと入力音声の子音開始部，終了部を視察によって定め，図6.8に示すようにDPマッチングにおける類似度和最大のパスから，子音開始部，子音終了部における標準パターンと入力の対応づけのずれを求める。発声者T，Iについて，サンプルaの30サ

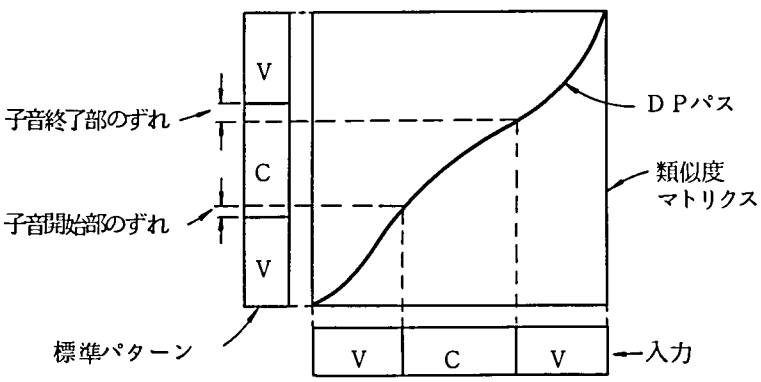


図 6.8　DPマッチングによる対応づけにおけるずれ

ンプルを用いてずれを求め，その分布を図6.9に示す。また，対応づけのずれの標準偏差を求めて表6.12に示す。図6.9から，ずれの分布は両者でほぼ等しいことがわかる。また，表6.12から，認識率の低い発声者Iの方がずれの標準偏差が小さいことがわかる。したがって，発声者Tについて入力と標準パターンの対応づけが正しくされていると考える限り，発声者Iについてもdpマッチングは正しく行われているといえる。

以上の考察から，DPマッチングによる時間軸の正規化は正しく行われており，認識誤りの原因とはいえないと結論できる。

(b)　類似度和を計算する際の重みづけの検討

6.4.2.3で提案した重みづけの方法の有効性について検討する。実験2の結果を発声者別の図にして，図6.10に示す。横軸は語頭，語尾の類似度を無視するフレームの数である。なお，図6.10は各発声者のＶＣＶ音節の平均の長さの1/3をあわせて示して

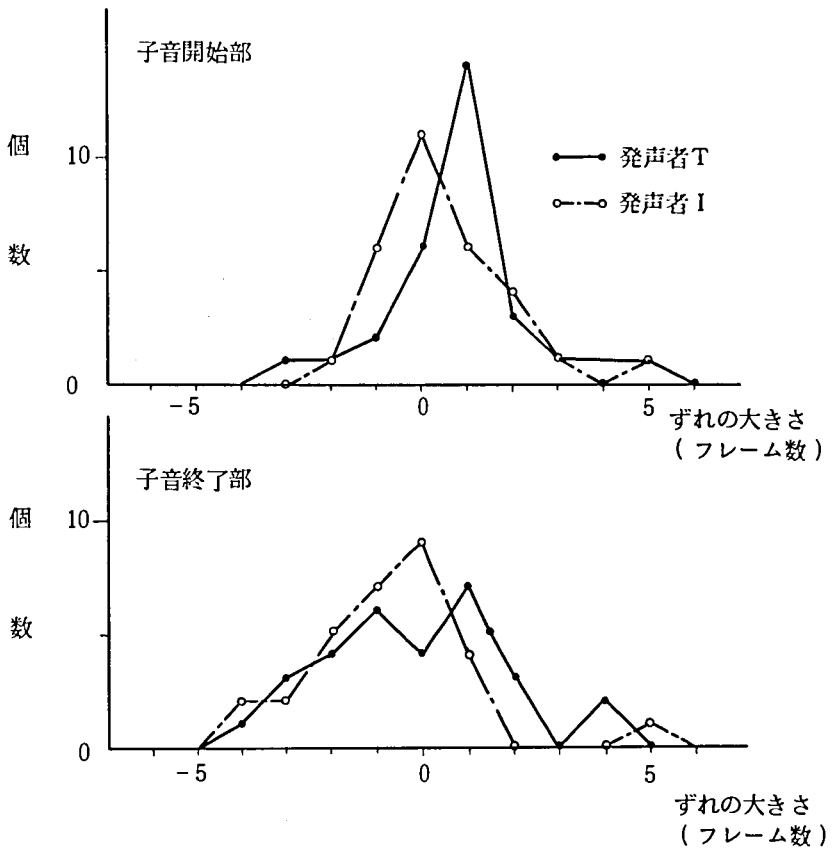


図 6.9　子音開始部，終了部におけるずれの分布

表 6.12　子音部のずれの分布の標準偏差

位置 / 発声者	子音開始部	子音終了部
T	1.53	1.96
I	1.41	1.73

ある。図 6.10 から次のことがわかる。すなわち，認識率の低い発声者は，前後の母音を 5 フレーム程度無視することにより認識率が上昇する。これに対し，認識率の高い発声者は，5 フレーム程度無視しても認識率はあまり変化しない。また，5 名の平均の認識率は，$k = 0$ のとき 80.3％，$k = 5$ のとき 84.7％であり，前後 5 フレーム無視することにより4.4％上昇する。また，図 6.10において，いずれの発声者も，無視するフレーム数がVCV音節の平均長の約 1/3 をこえると，急激に認識率が低下することが認められる。単独のVCV音節の中央

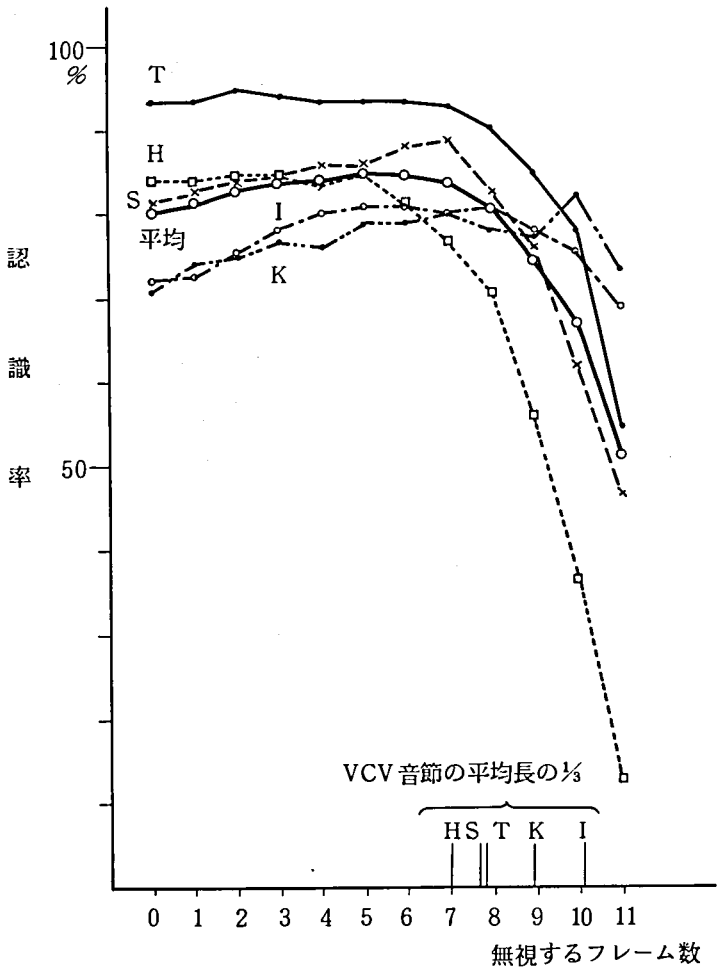


図 6.10　無視するフレーム数と認識率の関係

の約 1/3 の区間は，子音およびその近傍のわたりの区間と考えられるから，この区間が子音に関する情報を持っていることがわかる。これらの結果から，VCV音節の認識には子音およびその近傍の区間が重要であって，母音定常部は認識には不必要であり，発声者によってはむしろ認識に悪影響を及ぼすことがわかる。

次に，母音定常部がVCV音節の認識に及ぼす影響についてよりくわしく検討する。まず，次式で定義される曲線をえがいて，VCV音節のどの部分が認識誤りの原因になっているか調べる。

$$y_i = \sum_{j=1}^{i} l_j - \sum_{j=1}^{i} l_j^{*} \qquad (6.34)$$

（$i = 1, 2, \cdots\cdots, N$; NはVCV音節のフレーム数）

右辺第1項は，入力と同じカテゴリーに属する標準パターンとのDPマッチングにおける第iフレーム目までの類似度和であり，第2項は，それ以外のカテゴリーに属する標準パターンのうちで，同様の類似度和の最大値である。曲線y_iの性質から，認識誤りの原因を次の4つの種類に分類することができる。

a. 語頭の母音が認識誤りの原因　（図6.11）

b. 子音部分が認識誤りの原因　（図6.12）

c. 語尾の母音が認識誤りの原因　（図6.13）

d. VCV音節全体が認識誤りの原因（図6.14）

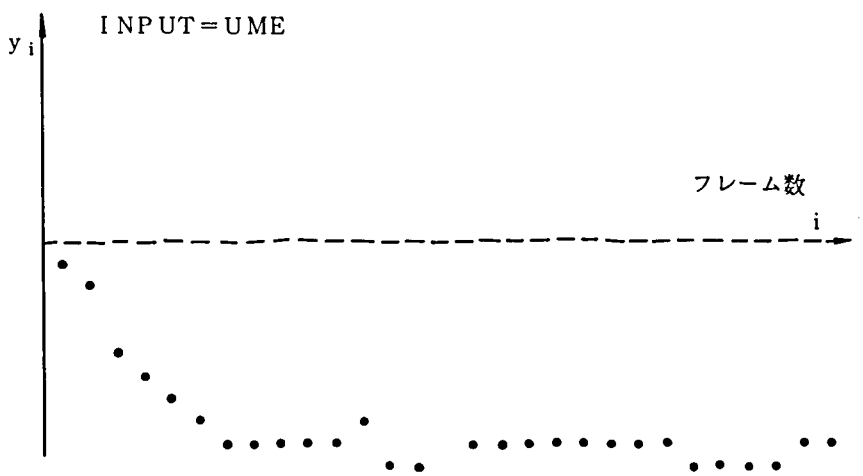


図6.11　語頭の母音が認識誤りの原因となっている例

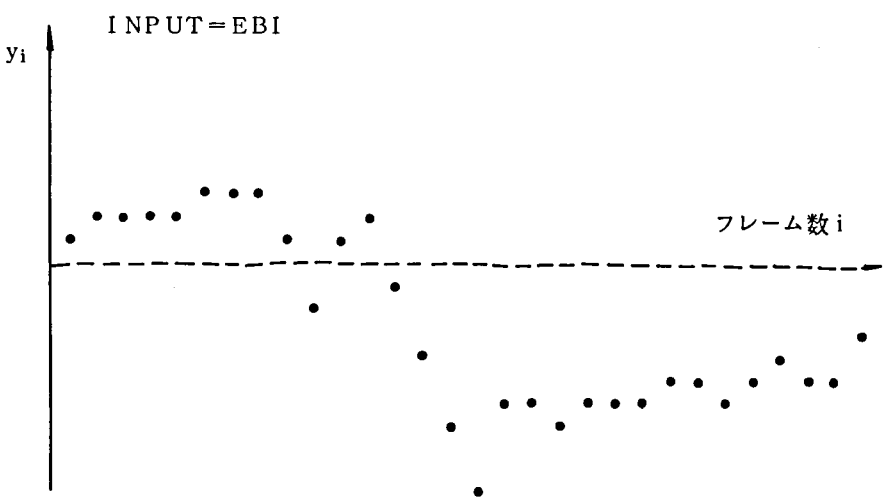


図6.12　子音が認識誤りの原因となっている例

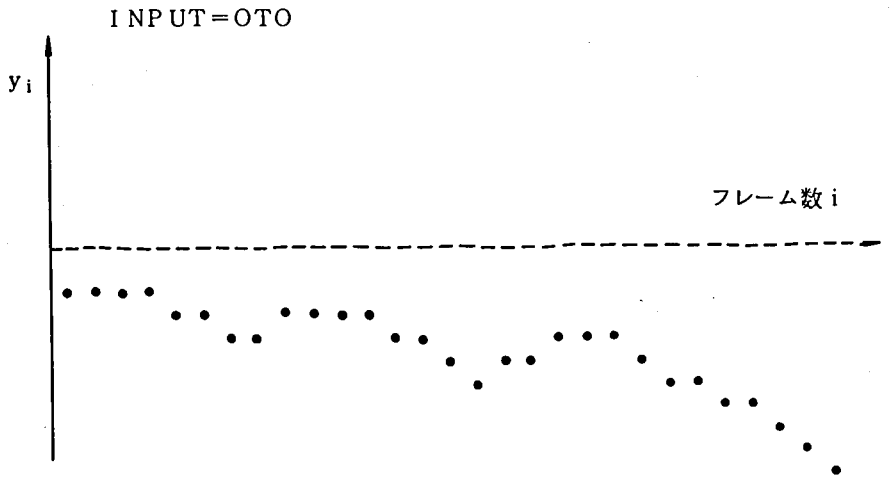


図 6.14　VCV音節全体が認識誤りの原因となっている例

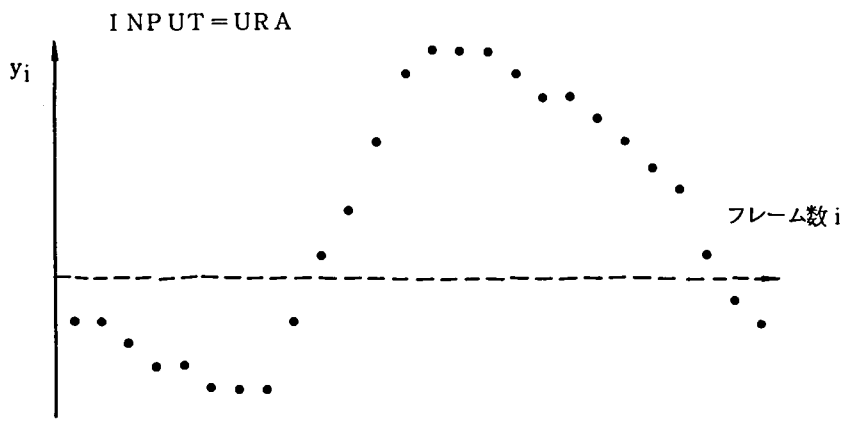


図 6.13　語尾の母音が認識誤りの原因となっている例

発声者T，Iについて，サンプルaを用いて認識した結果，生じた誤りを，原因別に分類したものを表 6.13に示す。表 6.13から，発声者Iの認識率が低いのは語頭，語尾（特に語尾）の母音が原因となっていることがわかる。これから，認識率の低い発声者は，母音のスペクトルの変動が大きいと予想される。次に，実際のサンプルについてこのことをたしかめる。

V_i（$i = 1, 2, 3, 4, 5$）……入力のVCV音節の母音部分の相関関数（5回の発声）　(6.35)

表 6.13　認識誤りの原因別の分類

発声者	a	b	c	d	計
T	5	4	29	4	42
I	1	4	5	0	10

a：語頭の母音が認識誤りの原因

b：子音が　　〃

c：語尾の母音が　　〃

d：VCV音節全体が　　〃

$\boldsymbol{V}_s$ …… 対応する標準パターンの母音部分の相関関数　　(6.36)

$\boldsymbol{M} = \frac{1}{5} \sum_{i=1}^{5} \boldsymbol{V}_i$ …… $\boldsymbol{V}_i$ の平均値　　(6.37)

$\boldsymbol{V}_i$ および $\boldsymbol{V}_s$ は，VCV音節から抽出した母音定常部の相関関数を平均したものである。これらのパラメータを用いて次の量を求める

$\boldsymbol{V}_i$ と $\boldsymbol{M}$ の平均距離

$\boldsymbol{V}_s$ と $\boldsymbol{M}$ の距離

ただし，距離の定義式としては最尤スペクトル分析法に基づくスペクトル間の差(*)を用いる。結果を表 6.14に示す。表 6.14から次のことがわかる。

(1)　発声者Kは入力の語頭，語尾の母音のばらつきが大きい。

＊　2章で2つのスペクトル間の距離を定義した（式(2.22)）が，この式は2つのスペクトルに関して非対称であった。そこで，ここでは次の式で2つの周波数スペクトル $T(w)$ と $R(w)$（w：角周波数）の距離を定義する。

$$\frac{1}{2}\left[\int_{-\pi}^{\pi}\left\{\log\frac{T(w)}{R(w)}+\frac{R(w)}{T(w)}-1\right\}dw+\int_{-\pi}^{\pi}\left\{\log\frac{R(w)}{T(w)}+\frac{T(w)}{R(w)}-1\right\}dw\right]$$

$$=\frac{1}{2}\int_{-\pi}^{\pi}\left\{\frac{R(w)}{T(w)}+\frac{T(w)}{R(w)}-2\right\}dw$$

表 6.14 母 音 の 変 動

発声者	P_i と M の平均距離		P_s と M の距離	
	語頭の母音	語尾の母音	語頭の母音	語尾の母音
T	1.09	1.08	1.11	1.97
I	1.25	1.38	1.53	3.43
K	2.16	1.74	1.70	1.90

(2) 発声者Iは，入力と標準パターンで語尾の母音の距離が大きい。

すなわち，発声者Kの認識率が低いのは，語頭，語尾の母音のスペクトルが，発声するたびに大きく変動しているためである。また，発声者Iの認識率が低いのは，入力と標準パターンとで語尾の母音のスペクトルが異なっているためである。このことは，発声者Iは語尾の母音が認識誤りの原因になっているという表 6.13の結果と良く合っている。

以上の検討から，VCV音節の母音定常部は発声者によっては変動が大きく，これがVCV音節の認識に悪影響を及ぼしていると結論できる。したがって，単独のVCV音節の認識においては語頭，語尾の母音定常部に相当する5フレーム程度を無視した方が良い。

(c) 線形マッチングの検討

自由度1，端点フリーという条件のDPマッチングは，6.4.2.2で述べたように線形のマッチングを行うことに相当する。この線形マッチングと通常のDPマッチングを比較検討する。線形マッチングを用いた実験（実験3）の結果を発声者別のグラフにして図 6.15に示す。横軸は図 6.10と同様に類似度和の計算の際無視するフレームの数である。図 6.15と図 6.10を比較すると，線形マッチングを用いた場合は，通常のDPマッチングの場合に比較して，認識率が上昇している発声者もいるが，全体では認識率が低下している。たとえば，$k = 5$ の場合，図 6.10では平均の認識率は 84.7％であるが，図 6.15では 79.1％に低下している。これは，子音部のマッチングにおいて，単なる重ね合わせよりDPによる時間軸正規化の方が有効であることを示している。すなわち，子音部およびその近傍の時間構造が非線形に変化していることを示している。また，時間構造の変化は個人により差があり，DPマッチングに比較して線形マッチングを用いた場合に認識率の低下のはげしい発声者K，Hは，時間構造が安定していないと考えられる。

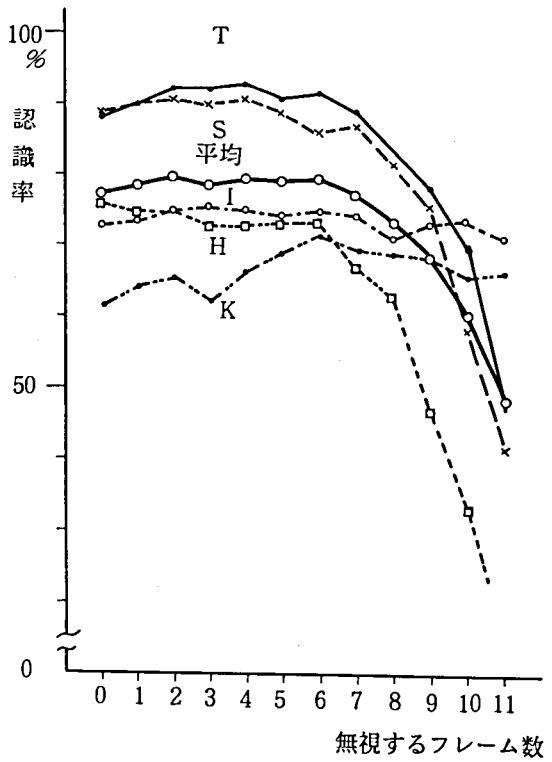


図 6.15　線形マッチングの結果

6.6.2.2　母音定常部を含まないVCV音節の認識法の検討

(a)　DPマッチングの検討

DPマッチング法を使うことの有効性を調べるため，6.6.2.1と同じようにVCV音節の時間構造を調べる。4名の発声者（T，I，H，K）の連続音声中から切り出されたVCV音節各30サンプルを用いて標準パターンとの長さの比を求め，その分布を調べた。結果を表6.15に示す。R_1～R_4は式（6.32）に示したのと同じものである。表6.15から次のことがわかる。

(1)　標準パターンと入力との時間構造は明らかに異なっているため，単に線形のマッチングでは正しい認識は行えない。したがって．DPによる時間軸正規化マッチングを用いるのが有効である。

(2)　子音部に比較して母音部のRが小さい。すなわち，単独のVCV音節に比べると，連続音声中のVCV音節は，母音部分の縮み方が大きい。これは，連続音声中では，母音が定常値（ターゲット）に到達する前に次の音韻に向かって動き出し，母音の定常部が存在しないためである。したがって，標準パターンと入力の端点を固定した従来のマッチングでは正しい対応づけ

表 6.15　入力と標準パターンの長さの比

発声者		R_1	R_2	R_3	R_4
T	平　均	0.567	0.797	0.736	0.644
	標準偏差	0.198	0.347	0.432	0.157
I	平　均	0.660	0.883	0.720	0.712
	標準偏差	0.239	0.337	0.419	0.198
H	平　均	0.581	0.697	0.402	0.505
	標準偏差	0.279	0.270	0.189	0.116
K	平　均	0.530	0.737	0.609	0.595
	標準偏差	0.187	0.261	0.278	0.138

が出来ないと考えられる。

(3)　自由度が3のマッチング法は，入力を最大2倍に引き伸ばす能力を持つ。しかしながら，表6.15を見ると，連続音声中のVCV音節の各部分は，標準パターンに比較して1/2以下に縮んでいる場合もあるため，自由度が3のDPマッチングでは正しい対応づけが行えない恐れがある。

(b)　端点条件，自由度の検討

(a)の考察の結果から，最適の端点条件，自由度を求めることにする。実験4は．端点固定，端点フリー，自由度3，自由度4の各条件を組み合わせた4種類の方法の比較実験である。結果を図6.16に示す。図6.16より次のことがわかる。

(1)　端点固定より，端点フリーの方が認識率が良い。これは，連続音声中のVCV音節は，単独のVCV音節の母音定常部を取り去ったものであるという仮定のうらづけになっている。

(2)　マッチングの自由度（パスの方向の自由度，端点固定かフリーかという自由度を含めて）を増すことはパターンの正規化の能力を増すことになるが，同時に，異なった音韻間のconfusionを増大させる原因ともなる。したがって，最適の自由度が存在すると考えられる。図6.16はこのことを示しており，端点フリー，自由度3の場合が最適である。

(3)　発声者Hのみは端点フリー，自由度4の方法が最も認識率が良い。これは，表6.15か

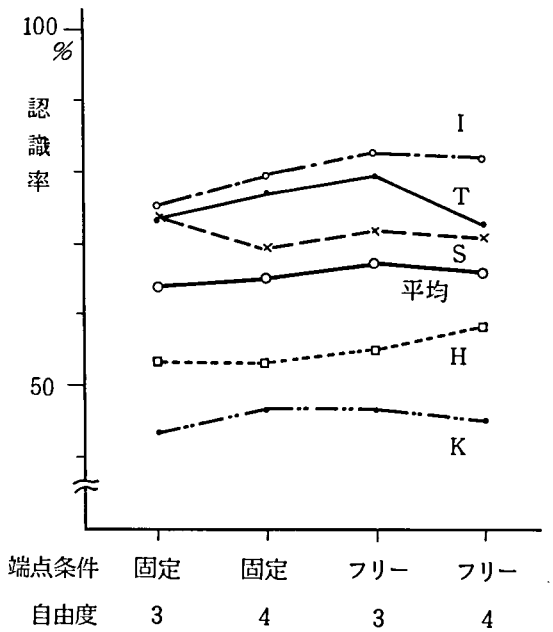


図 6.16　端点条件，自由度を組み合わせた比較実験

らわかるように，発声者Hが標準パターンに比べ入力の子音の縮少率が最も大きく，自由度が3のDPでは正しい対応づけができない場合が多いからである。図 6.17にDPパスの例を示しておく。

次に，もっと多くのデータについて端点固定，端点フリーの比較を行う。サンプルd～fを使った認識実験（実験1，実験5）の結果を図 6.18に示す。図 6.18から次のことがわかる。

(1)　端点固定の場合，平均の認識率は 60.1％であるが，端点フリーにすると 66.0％と約6％上昇し，端点フリーの方が有効であることがわかる。

(2)　しかも，端点固定の場合は音節数の増加と共に認識率が低下する傾向がみられたが，端点フリーの場合にはそのような傾向がなくなり，音節数が増加してもほぼ一定の認識率が得られるようになった。これは，長い連続音声の認識を行おうとする場合に好ましい結果である。

(c)　類似度和の重みづけの検討

6. 6. 1. 1で述べたように，母音定常部を含んだVCV音節の場合は，前後の数フレームの類似度を無視することが有効であった。連続音声中のVCV音節に対して同じ方法が有効かどうか検

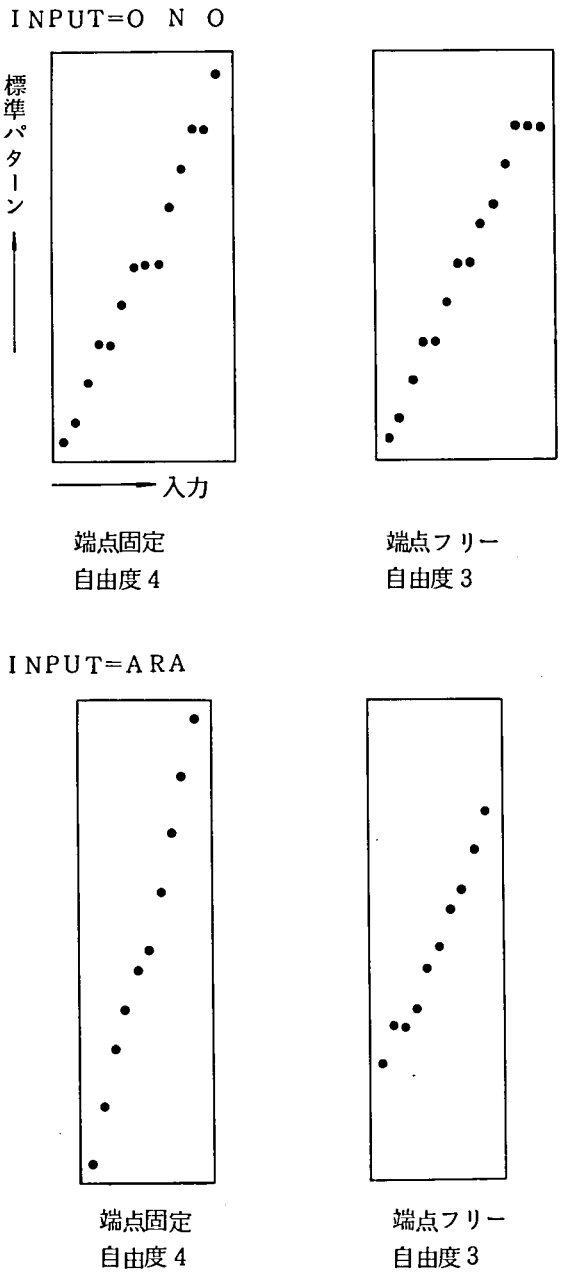


図 6.17　DPマッチングにおけるパスの例

討する。実験5の結果を図6.19に示す。横軸は，語頭，語尾の無視するフレーム数である。図6.19を見ると，図6.10のように5フレームまで無視すると認識率が上昇するという現象が見られない。2フレーム程度までは安定であるか，もしくは，発声者によっては少々認識率が上昇するが，それ以上のフレームを無視すると急激に認識率が低下する。したがって，連続音声中のVCV音節のように，母音定常部を含まないVCV音節の認識では，重みづけを行わない方が

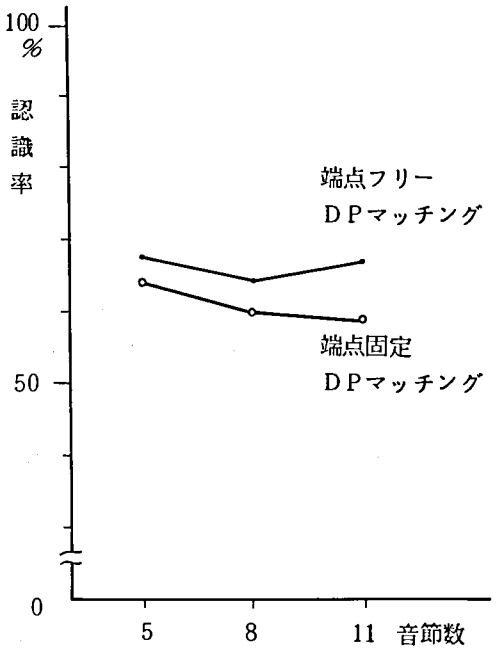


図 6.18　端点固定，端点フリーDPマッチングの比較

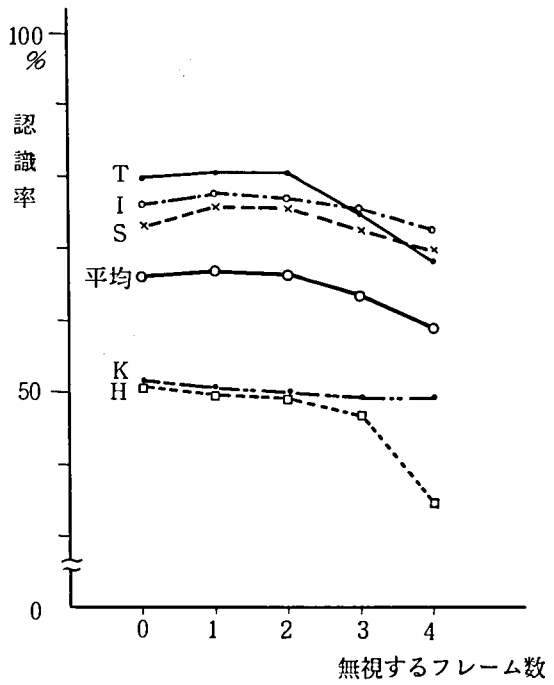


図 6.19　無視するフレーム数と認識率の関係
（サンプルd〜f　端点フリーDPマッチング）

良いといえる。

6.6.3 結　論

以上の検討事項をまとめると次のようになる。

(1) 標準パターンに関して：連続音声中のVCV音節を認識する際は，単独のVCV音節から作成した標準パターンでも，連続音声中のVCV音節から作成した標準パターンでも，同じ程度の認識率が得られる。標準パターン作成の手間を考えると，単独のVCV音節から作成した標準パターンの方が有利である。

(2) 母音定常部を含んだVCV音節の認識に関して：母音定常部は変動が大きく，認識に悪影響を及ぼすため，無視した方が良い。無視するのは，VCV音節の語頭，語尾の5フレームが適当である。

(3) 母音定常部を含まないVCV音節の認識に関して：単独のVCV音節から作成した標準パターンとマッチングする際は，入力を標準パターンの一部に対応づける端点フリーDPマッチングが有効である。

6.7 端点フリーDPマッチングを用いた連続単語音声の認識[24][119][124]

6.6では，連続音声中のVCV音節を認識するには，自由度3，端点フリーのDPマッチング（以後これを端点フリーDPマッチングと呼ぶ）が有効であることがわかった。その応用として，ここでは，前章で述べた連続単語音声認識に端点フリーDPマッチング法を適用した結果を述べる。

6.7.1 連続単語音声認識系の構成

図6.20に連続単語音声認識系の構成を示す。詳細については第5章に述べてあるので，ここでは各部について簡単に述べる。

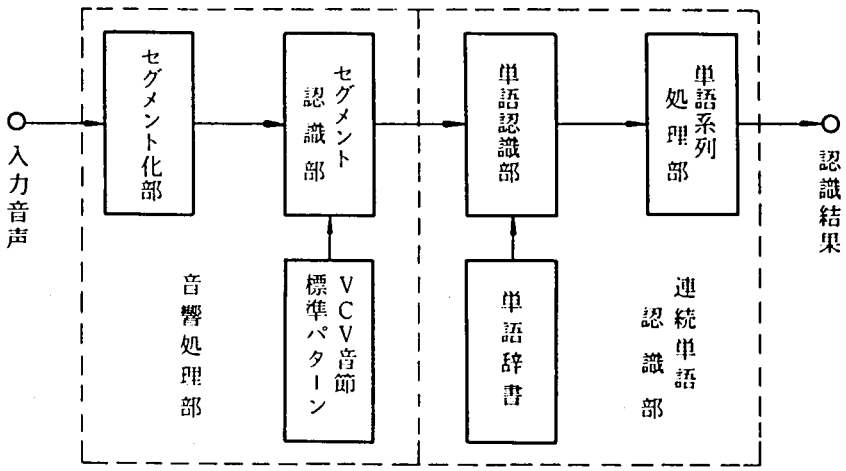


図 6.20　連続単語認識系の構成

(1)　音響処理部

入力音声を，まずセグメント化部でVCV音節単位にセグメンテーションし，かつ，母音認識を行う。セグメンテーションの際，一意に決定するのが困難であれば，あいまいさを残しておく。次に，セグメント認識部でVCV音節標準パターンとのDPマッチングを行い，VCV音節単位の子音認識を行う。処理結果は，セグメンテーションのあいまいさ，および，類似度2位以下の結果を含めて，図 6.21に示すVCVマトリクスの形にして次段に送る。

セグメント番号 →

類似度順位 ↓

	1	2	3	4	5	6
1	IBU 422	UKU 496	URO 273	OMA 274	AZI 408	
2	IBO 421	UKO 463	ORO 262	OA 265	AU 201	UZI 198
3	IMO 413	UPU 457	UGO 257	OGA 259	ODI 388	
4	IGU 408	UPO 448	UBO 250			
5	IGO 407	OKU 447	UNO 250			

図 6.21　VCVマトリクス

(2) 連続単語認識部

単語認識部では，単語の始点，終点（VCVマトリクス上の第i列，第j列とする）が指定されたとき，その区間における最もたしかな単語$w(i, j)$，および，その類似度$L(i, j)$を求める。単語系列処理部では，単語認識部の結果を用いて，音声区間全体での類似度和を最大にするという評価基準のもとで，最適の単語系列を求める。すなわち，

$$L_m = \max_{1 \leqq i_1 < i_2 < \cdots < i_n < m} \{ L(1, i_1) + L(i_1+1, i_2) + \cdots \cdots + L(i_n+1, m) \} \} \quad (6.38)$$

（mはVCV音節ラティスの列の数）

を満足する単語系列

$$w(1, i_1)\, w(i_1+1, i_2) \cdots\cdots w(i_n+1, m) \quad (6.39)$$

が連続単語の認識結果であるとする。式（6.38）は次のように漸化式表現になおすことができる。

$$L_0 = 0, \quad L_j = \max_{i=1}^{j} \{ L_{i-1} + L(i, j) \} \quad (6.40)$$

したがって，DPを用いて簡単に計算できる。

6.7.2 認識実験

認識対象は，（6.4）～（6.6）に示した連続単語である。3名の発声者（T，I，S）が各連続単語を2回ずつ発声した計540サンプルを認識実験に用いた。VCV音節単位のセグメントの認識には，端点固定DPマッチング，端点フリーDPマッチングを用いた2種類の実験を行った。結果を表6.16に示す。

表6.16 連続単語中の単語の認識率

セグメント認識法＼発声者	T	I	S	平均
端点固定DPマッチング	90.7％	92.0％	93.7％	92.2％
端点フリーDPマッチング	93.3％	96.1％	95.0％	94.8％

6.7.3 検　討

表 6.16によれば，端点固定DPマッチングを用いた場合に比較して，端点フリーDPマッチングを用いると，認識率が 2.6％向上し，94.8％の認識率が得られた。次に認識誤りの原因について調べる。認識誤りは，(1)セグメント化，(2)母音認識，(3)VCV音節セグメントの認識，のいずれかが原因で生じる。認識誤りを原因別に分類して図 6.22に示す。端点固定DPマッチングの場合は(3)の原因で生じた誤りが最も多かったのに対し，端点フリーDPマッチングを用いると，そのような誤りが大幅に減少したことがわかる。以上のように，実際の認識に応用した場合にも，端点フリーDPマッチングが有効である。

6.8　あとがき

以上述べてきたように，本章では，VCV音節認識の改良を目的として，種々の標準パターン作成法，認識法を提案し，その比較検討を行った。その結果，VCV音節標準パターンに関して

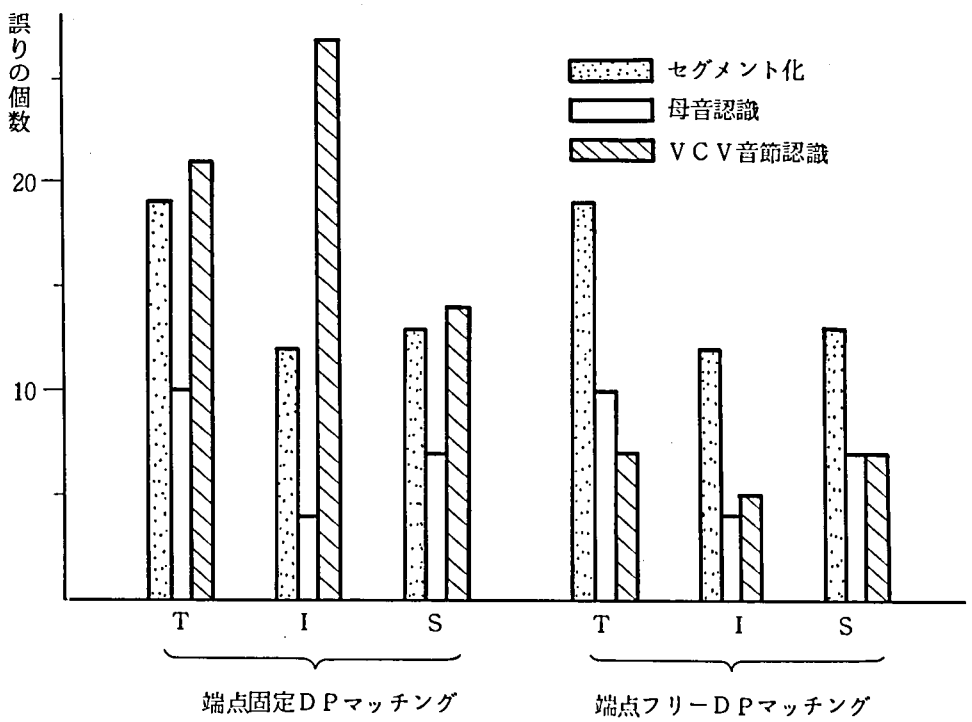


図 6.22　認識誤りの原因別分類
T，I，S ……発声者

は，連続音声中のVCV音節から作成する必要はなく，単独に発声したVCV音節より作成した標準パターンで十分であることを示した。VCV音節の認識法に関しては，問題を2つの場合に分けて論じた。まず，単独のVCV音節のように，母音定常部を含むVCV音節の認識の際は，類似度和を計算する時，前後数フレームを無視するのが有効であることを示した。認識実験の結果，無視するフレーム数は5フレームが適当であり，またその場合，認識率が4.4%向上することがわかった。次に，連続音声中のVCV音節のように母音定常部を含まないVCV音節の認識には，パスの自由度が3で端点フリーのDPマッチングが有効であることを示した。また，認識実験の結果，この方法を用いると端点固定の方法に比較して認識率が5.9%向上することがわかった。最後に，端点フリーDPマッチング法を連続単語音声の認識に適用した結果，認識率が2.6%向上して94.8%の認識率が得られた。

本章で提案した端点フリーDPマッチングは，VCV音節の認識に有効であるばかりでなく，他にも多くの応用が可能である。すなわち，この方法は一方のパターンを他方のパターンの一部にマッチングさせる方法であるから，連続音声中の単語と，単語の標準パターンとのマッチングが容易に行える。したがって，連続音声中から特定の単語を検出するいわゆるword spotting や，その応用としての連続単語音声認識に適用可能である。同じような考え方に基づいたword spotting についてはいくつかの試みが行われている。[125][126] また，端点フリーDPマッチングに基づいた連続単語音声認識については第9章で述べる。

第7章　日本語会話音声認識システムの検討（第1次システム）

7.1　はしがき

前章までVCV音節を単位とした単語音声の認識について述べた。次の段階として，人間にとって自然な形式である会話音声を認識対象とした会話音声認識を取扱う。会話音声においては，単語が連続して発声されるだけでなく，構文，意味といった高次のいわゆる言語情報を含んでいるために，その取扱いは単語音声，連続単語音声に比較すると非常に困難である。会話音声認識に関する初期の研究は，言語情報の取扱いをさけ，音声をそのまま音韻の系列に変換しようとするいわゆる音声タイプライタの研究である。[75][76]その結果，言語情報を使うことの必要性が認識され，また，最近さかんになった自然言語処理研究に刺激され，言語情報を積極的に使って会話音声を認識しようとする動きが生じてきた。これは音声理解系（Speech Understanding System）の研究と呼ばれており，アメリカでは1971年にARPAの援助のもとに大がかりな音声理解系のプロジェクトが開始された。[8]その後，我が国やフランス，イタリア等でも同様の研究が開始された。[80]〜[97]音声理解系の基本的な考え方は次の3点である。

(1)　会話の内容を限定していること。

(2)　システムは，発話された個々の音韻や単語の認識より，音声に含まれる意味内容を理解することに重点をおいていること。

(3)　質問回答形式で発声者との対話を行う形式をとっていること。

著者は，前章までに述べてきた単語音声認識，連続単語音声認識の成果の上に立って，日本語会話音声の認識システムの研究を1974年に開始した。ここでは会話の対象として列車の座席予約サービスをとり，質問回答形式で入力音声を理解するシステムの作成を目ざした。第1次システムは1975年に完成し，さらにいくつかの改良を加え，1976年に第2次システムを完成した。本章では，音響処理を中心に第1次システムの内容について述べる。第2次システムについては次章で述べる。

7.2 第1次システムの概要 (127)(129)(130)

会話音声認識の研究では，研究対象の選定が重要な問題である。研究対象の選定にあたっては，次の6つの点を考慮した。

(1) 会話の話題が限定されている。

(2) 入力の語彙は数十～数百程度である。

(3) 入力の文章は簡単な構文構造である。

(4) 入力の文章に意味上の冗長性がある。

(5) 会話のやりとりがある程度継続する。

(6) 発声者ごとに個人性の特徴を学習することができる。

研究対象の候補として，音声ダイヤル，計算サービス，座席予約サービス，番号案内サービス，情報検索，音声タイプライタなどを取り上げて，上の各事項について適合性を比較検討した。その結果を表7.1に示す。この結果から座席予約サービスが適当であると判断し，研究対

表7.1 研究対象の適合性の比較

適合事項 / 研究対象	話題が限定されているか 否か	入力の語彙が少なくてすむか 否か	入力の構文が簡潔か 否か	入力に冗長性があるか 否か	会話の継続性があるか 否か	個人性の学習が行いやすいか 否か
音声ダイヤル	◎	◎	◎	×	×	×
計算サービス	○	◎	◎	×	×	△
座席予約サービス	○	○	○	○	○	△
番号案内サービス	○	△	○	○	△	△
情報検索	○	△	○	○	○	△
音声タイプライタ	×	×	×	×	×	○

注) ◎○△×の順に適合事項に合致しているものとする。

象として取り上げた。

認識対象の内容を表7.2に示す。対象は新幹線の座席予約サービスであり，質問回答をくり返すことによって発声者が意図する予約内容を認識する形式になっている。予約項目は，発駅，

表7.2 認識対象

タスク	列車の座席予約（新幹線）
単語数	65（駅名13）
予約項目	発駅，着駅，発時刻，列車名，等，枚数
発声の制限	文節ごとにポーズをおく
発声例	新大阪から，名古屋までの，ひかり26号のグリーンを，3枚，お願い致します。

着駅，発時刻，列車名，等，枚数の6項目であり，入力の文章は予約項目の文節ごとに区切って発声する。認識の対象となる単語は，表7.3に示すように13個の駅名，その他の名詞，助詞，動詞等を加えて，合計65種類である。入力音声に含まれる音韻を表7.4に示す。母音は5種類，子音は18種類で，計23種類である。撥音／N／は本システムでは母音と同様に扱う。

表7.3 認識対象の単語

名詞	駅名	東京 新横浜 小田原 熱海 三島 静岡 浜松 豊橋 名古屋 岐阜羽島 米原 京都 新大阪
	数字	0 1 2 3 4 5 6 7（なな，しち） 8 9 10 100 ひとり ふたり
	その他	こだま ひかり 普通 グリーン 時 分 駅 枚 人 号 指定 券 席
助詞		から より 発 の で を のを まで へ いき ゆき は が に のに
動詞		予約 お願い いた し ます あり か
その他		はい いいえ

表 7.4　音声中に含まれる音韻

母 音	a, i, u, e, o,
子 音	N, m, n, b, d, g, r, z, p, t, k, s, h w, y, zy, ky, hy

会話音声認識システム全体の構成を図 7.1 に示す。システムは，音響処理部，言語処理部，音声応答部の 3 つの部分から構成されている。音響処理部と言語処理部で入力音声の認識を行い，認識された予約内容を音声応答部に渡す。音声応答部では，予約の問い合わせや確認を行

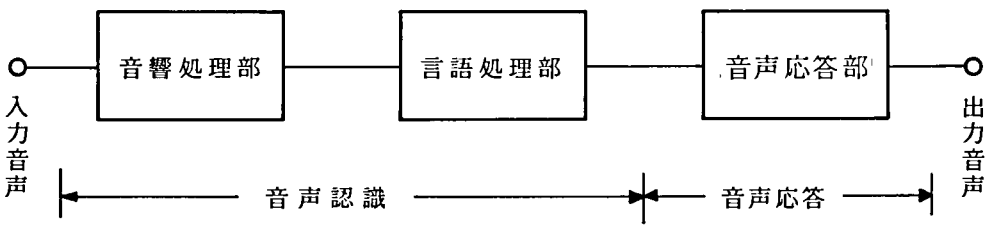


図 7.1　会話音声認識システムの構成

い，それを合成音で発声者に伝える。このようにして，システムと発声者の間で質問回答をくり返すことによって正しい予約内容がシステムに認識される。表 7.5 に質問回答の例を示す。

表 7.5　列車の座席予約に関する質問回答の例
（Q：音声応答，A：入力音声）

Q：こちらは電話予約サービスです。
　ご希望をどうぞおっしゃって下さい。
A：新大阪から，名古屋まで，ひかり 26 号を，お願い致します。
Q：どこまでですか。
A：名古屋までを，6 枚，下さい。
Q：普通券ですか。
A：はい。
Q：あなたの予約は，ひかり 26 号，新大阪駅 9 時55分発，名古屋までの普通券を 5 枚ですね。
A：いいえ，6 枚です。
⋮

7.3 音響処理 (128)

音響処理部の構成を図7.2に示す。入力音声は前処理部で特徴抽出された後，セグメント化部でVCV音節単位に分割され，次にセグメント認識部でVCV音節標準パターンを用いてセグメントの認識が行われる。最後にこれらの結果を音韻やセグメントのあいまいさを含んだ音韻ラティスの形で表現して出力する。以下，各部の詳細について述べる。

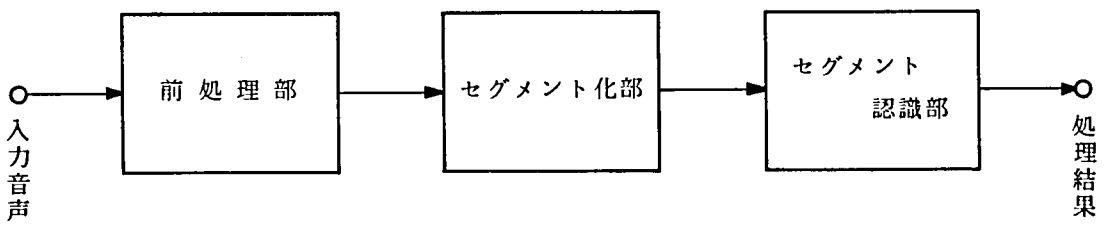


図7.2 音響処理系の構成

7.3.1 前処理部

前処理部の構成を図7.3に示す。この構成は第2章で述べた音声分析系と基本的には同一であるが，音声区間抽出法が異なっている。

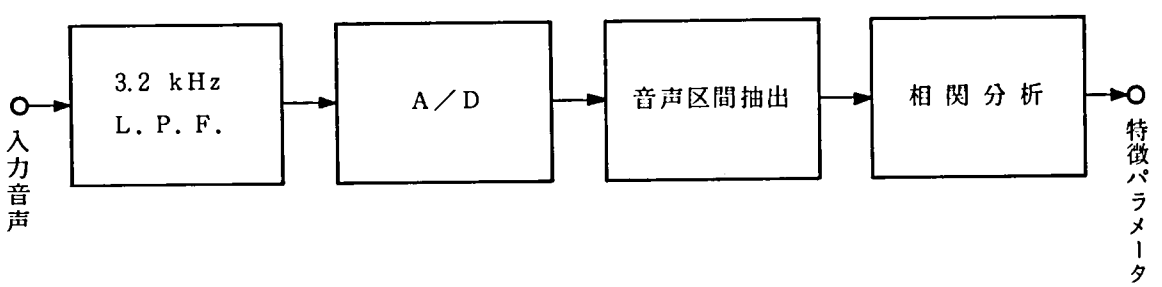


図7.3 前処理部の構成

入力音声は3.2kHz低域通過フイルタを通した後，標本化周波数8kHzのAD変換器で11ビットのディジタル音声に変換される。ディジタル音声は15msecごとに区分され音声パワーが求められる。文頭の子音を検出するため次のようにして音声区間の始端を検出する。2種類の閾

値P_1，P_2（$P_1 > P_2$）を定める。P_1より大きい音声パワーが検出されるとその10フレーム前から再び分析し，P_2よりパワーが大きくなった時点を始端とする。また，P_1以下のパワーが30フレーム（450msec）以上続いたとき，最初に閾値以下になったフレームを音声区間の終端とする。この手順を図7.4に示す。抽出された音声区間では音声の特徴量として15msecごとに波

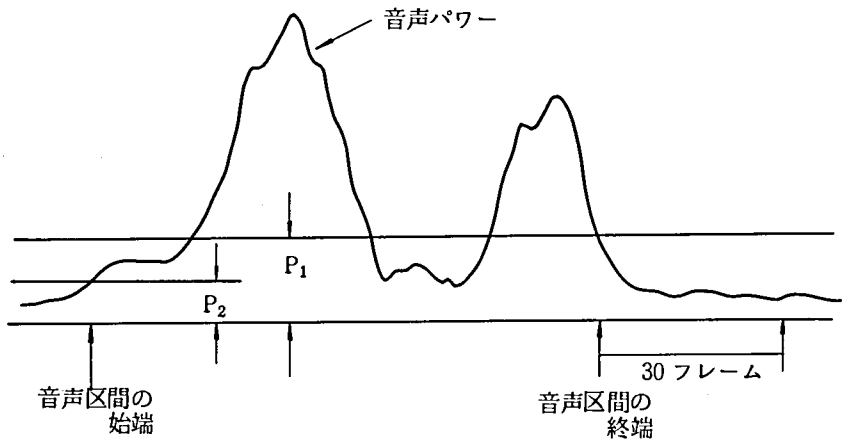


図7.4 音 声 区 間 の 抽 出

形の自己相関関数

$$\boldsymbol{v} = (v_0,\ v_1,\ \cdots\cdots\cdots\cdots,\ v_p) \qquad (7.1)$$

が求められる。ただしv_0は音声パワーである。音声区間の長さをNフレームとすると，次段へ送られる情報は自己相関関数の時系列

$$\boldsymbol{V} = (\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots\cdots\cdots,\ \boldsymbol{v}_i,\ \cdots\cdots\cdots,\ \boldsymbol{v}_N) \qquad (7.2)$$

である。ただし，$\boldsymbol{v}_i$は第iフレームの自己相関関数であり

$$\boldsymbol{v}_i = (v_{i0},\ v_{i1},\ \cdots\cdots\cdots\cdots,\ v_{ip}) \qquad (7.3)$$

と表現される。特に，v_{i0}は第iフレームの音声パワーである。

7.3.2 セグメント化部

セグメント化部の構成を図7.5に示す。

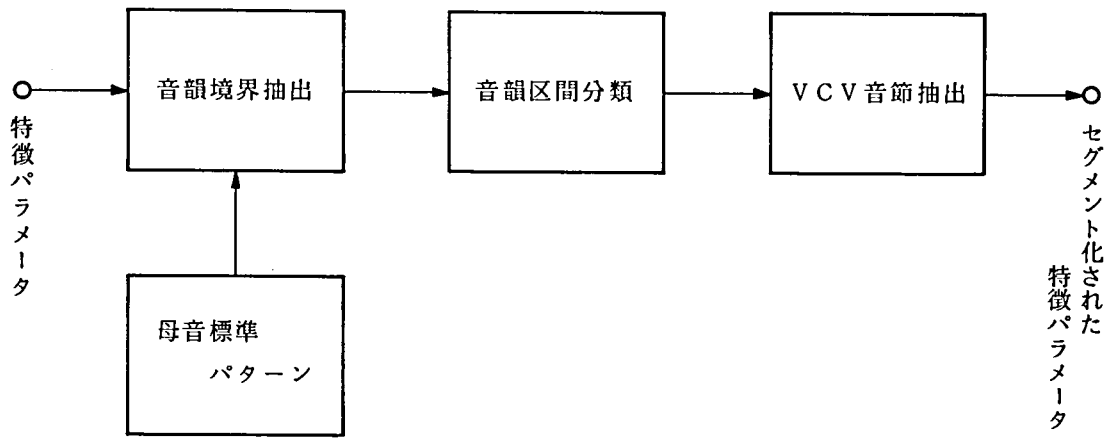


図7.5 セグメント化部の構成

7.3.2.1 音韻境界の抽出

入力音声の音韻境界抽出のために用いられる情報は次の4種類である。

(1) 音声パワーの系列

$$(v_{10},\ v_{20},\ \cdots\cdots\cdots\cdots,\ v_{N0}) \qquad (7.4)$$

(2) 急激なスペクトルの変化…第iフレームと第jフレームのスペクトルの差をdij*とすると，次の系列を急激なスペクトルの変化をあらわす量として用いる。

$$(e_2^1,\ e_3^1,\ \cdots\cdots\cdots\cdots,\ e_{N-2}^1) \qquad (7.5)$$

(ただし $e_i^1 = (d_{i-1,i+1} + d_{i,i+2})/2$)

(3) ゆるやかなスペクトルの変化…(2)よりも時間的にゆっくりとしたスペクトルの遷移を検出するために次の系列を用いる。

$$(e_3^2,\ e_4^2,\ \cdots\cdots\cdots\cdots,\ e_{N-3}^2) \qquad (7.6)$$

(ただし $e_i^2 = (d_{i-2,i+2} + d_{i-1,i+3})/2$)

---

* 第5章，第6章参照

(4) 母音系列…母音標準パターンを用いて各フレームの認識を行い，得られた類似度第1位および第2位の母音系列を用いる。類似度計算は次の手順で行われる

母音標準パターンは，最尤スペクトルパラメータの形でたくわえられている。母音/x/の標準パターンを

$$(A_{x0},\ A_{x1},\ \cdots\cdots,\ A_{xp}) \qquad (7.7)$$

とすると，第iフレームの自己相関関数と/x/との類似度は

$$l\ (i,\ x) = -\log\ (\sum_{\tau=0}^{p} A_{x\tau}\ v_{i\tau}) \qquad (7.8)$$

で与えられる。

以上の情報を用いて次の手順で音韻境界を抽出する。

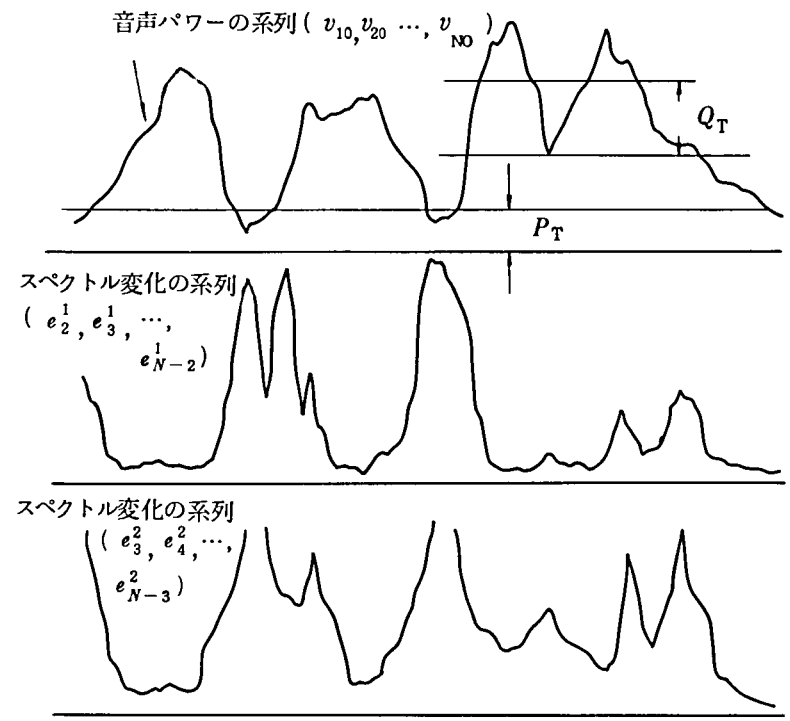
(1) 子音の多くは前後の母音に比較して口のせばめ等のためパワーが小さくなる。この性質を利用して音韻境界の抽出を行う。2種類の閾値P_T, Q_Tを定める。まず，パワーがP_Tより小さくなる区間を音韻境界とする。次に残った区間でパワーの極小点を検出する。前後の極大点との差が共に閾値Q_Tより大きい極小値を与える点を音韻境界とする。以上の手順で抽出された区間を音韻境界Ⅰと呼ぶ。

(2) 音韻の境界ではスペクトルが大きく変化するため，式（7.5），（7.6）の系列を使って音韻境界を抽出する。式（7.5），（7.6）が閾値以上である点，および前後の極小点との差が閾値以上である極大値を与える点を音韻境界とする。これらは，式（7.5），（7.6）の系列に負符号をつけたものに手順(1)と同じ方法を適用することによって求められる。この結果抽出されたものを音韻境界Ⅱと呼ぶ。

(3) 誤りを防ぐため音韻境界Ⅰの前後数フレーム以内にある音韻境界Ⅱを削除する。次に音韻境界Ⅰと音韻境界Ⅱをまとめて（すなわちORをとることにより）新たな音韻境界とする。

(4) 以上の手順でまだ抽出されない音韻境界を母音系列を用いて抽出する。同じ母音記号が類似度第1位の母音系列中に4個以上連続して現われる場合は，その母音が安定に存在すると考え，これを安定母音と呼ぶ。(1)～(3)の手順で決定した音韻区間のうち，長さがある閾値以上のものに注目する。その音韻区間中に安定母音が2個以上存在する場合は，それらの境界を音韻境界とする。

以上の手順で音韻境界が抽出される。抽出例を図7.6に示す。



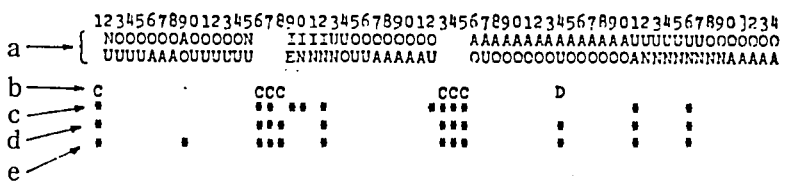


図 7.6　音 韻 境 界 抽 出 手 順

a. 第 1 位，第 2 位の母音系列

b. 手順(1)の結果（音韻境界 I ）

c. 手順(2)の結果（音韻境界 II ）

d. 手順(3)の結果

e. 手順(4)の結果（最終的な結果）

7.3.2.2　音韻区間の分類

抽出された音韻区間を母音，子音に分類する。分類のために用いる情報は次の3種類である。

(1)　各区間の継続時間

(2)　各区間の母音認識結果…次のようにして各区間の母音認識を行い，その結果を分類のための情報として用いる。

母音/x/が区間内の類似度第1位，および第2位の母音系列に現われる頻度をそれぞれ$N_1(x)$，$N_2(x)$とする。このとき，次の量$N(x)$を定義する。

$$N(x) = 4\,N_1(x) + N_2(x) \qquad (7.9)$$

$N(x)$の値が上位のものから2個を母音の候補としてえらぶ。また，区間内に安定な母音が存在する場合はその母音に一意に決定する。

(3)　音韻区間の両端の音韻境界の性質…(1)，(2)の補助的な情報として，音韻区間の両端の音韻境界がパワー系列によるものか，スペクトル系列によるものか等の情報を用いる。

以上の情報を用いて図7.7に示した手順で分類を行う。分類が困難である場合は未定のま

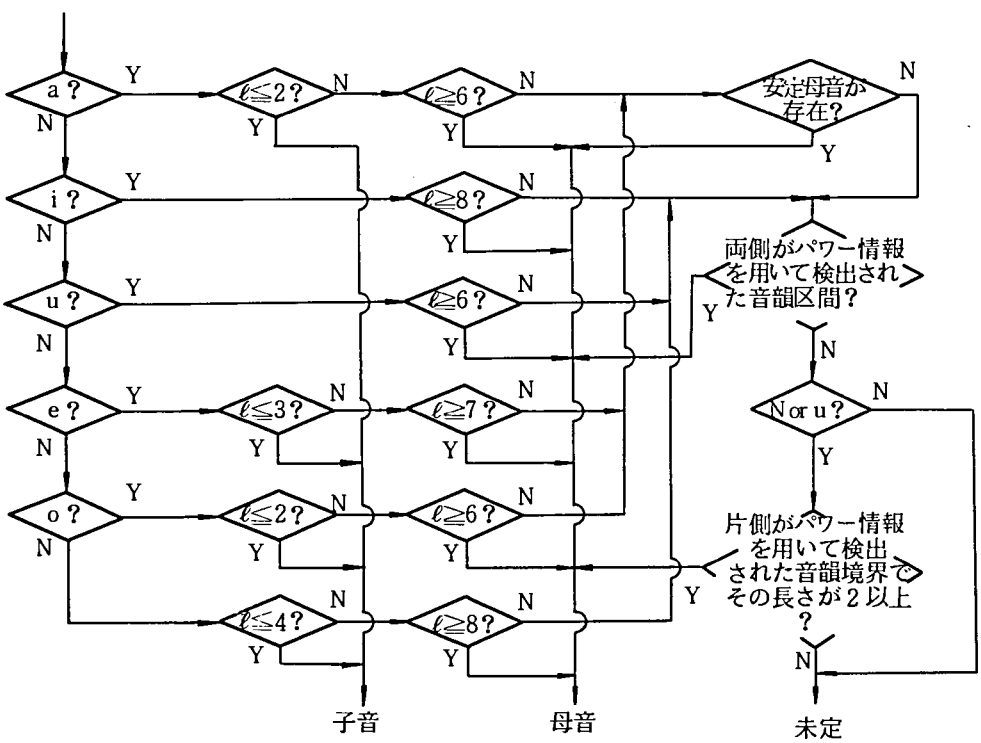


図7.7　音韻区間の分類（ℓ: 区間のフレーム数）

ま残しておく。

7.3.2.3 ＶＣＶ音節の抽出

7.3.2.2の手順で母音と判定された区間の中央を新たにセグメントの境界とすることにより．ＶＣＶ音節セグメントが抽出される。未定のまま残されたセグメントは，まだ母音，子音の判定がされていないから，抽出は一意に定まらず，２通り以上の抽出法が存在する。図7.6と同じ例について，ＶＣＶ音節を抽出した結果を図7.8に示す。このように，セグメントのあいまいさ

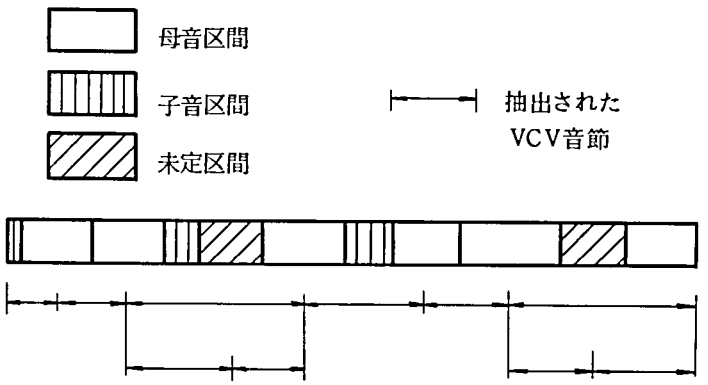


図7.8　ＶＣＶ音節の抽出

を残すことにより，セグメント化の段階での誤りを防ぐことができる。

以上の手続きを経て入力音声はＶＣＶ音節単位にセグメント化される。セグメント認識部へは次の情報が送られる。（図7.9）

(1)　相関関数の時系列（式（7.2））

(2)　各ＶＣＶ音節の先頭と末尾のフレーム番号（図7.9　A_1, A_2）

(2)　各ＶＣＶ音節の先頭と末尾の母音の候補（図7.9　B_1, B_2）

A_1	A_2	B_1	B_2
1	5		O
6	12	O	O
13	20	O	I N
13	27	O	O
21	27	I N	O
28	39	O	A
40	47	A	A
48	53	A	U
48	64	A	O
54	64	U	O

図7.9　認識部へ送られる情報

7.3.3　セグメント認識部

入力音声から切り出されたVCV音節セグメントと，VCV音節標準パターンとのDPを用いた時間軸正規化マッチングにより，セグメントの認識を行う。

(1)　相関関数の時系列として表現されている音声区間からVCV音節に相当する部分を切り出す。切り出されたVCV音節の1つを

$$(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\cdots, \boldsymbol{v}_V) \qquad (7.10)$$

(ただし $\boldsymbol{v}_i = (v_{i0}, v_{i1}, \cdots\cdots, v_{ip})$)

とする。

(2)　VCV音節標準パターンは最尤スペクトルパラメータの時系列として蓄えられている。切り出されたVCV音節の母音部分はすでに候補が決っているから，母音部分の一致する標準パターンのみとマッチングを行う。マッチングすべき標準パターンの1つを

$$(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots\cdots, \boldsymbol{A}_M) \qquad (7.11)$$

(ただし $\boldsymbol{A}j = (A_{j0}, A_{j1}, \cdots\cdots, A_{jp})$)

とする。

(3) 式（7.10），（7.11）から類似度マトリクスLMを作成する。

$$LM=\left\{ l(i,\ j)\right\} \qquad (7.12)$$

$$(i=1,\ 2,\ \cdots\cdots,\ \boldsymbol{v}\ ;\ j=1,\ 2,\ \cdots\cdots,\ M)$$

ただし，

$$l\ (i,\ j)=-\log\ (\sum_{\tau=1}^{p} A_{j\tau}\ v_{i\tau}) \qquad (7.13)$$

である。

(4) 類似度マトリクス上で，類似度和が最大のパスを探索する。すなわち，

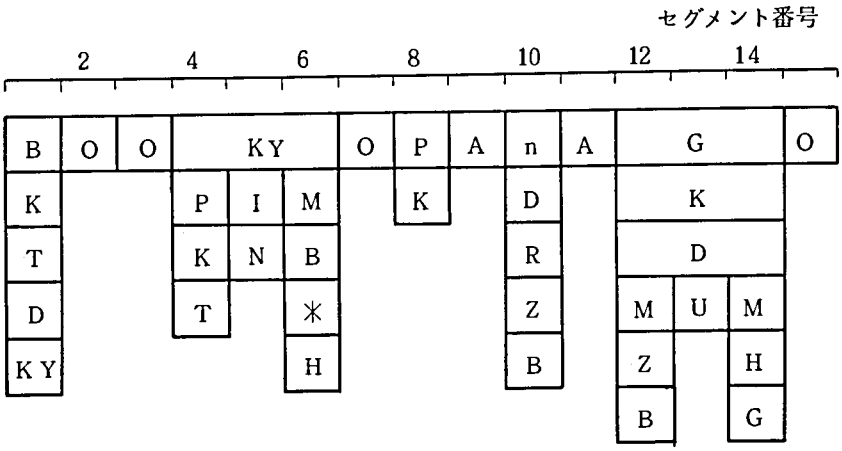
$$L=\max_{f}\left\{ \sum_{i=1}^{v} l\ (i,\ f(i)\ )\right\} \qquad (7.14)$$

ただし，

$$f(i)=\begin{cases} f\ (i-1) \\ f\ (i-1)+1 \\ f\ (i-1)+2 \end{cases} \qquad (7.15)$$

を満足するLを求めれば，これが式（7.10），（7.11）を時間軸の非線形正規化を行って得られた類似度である。なお，文頭にはCV型の音節が現われるが，これはCV型の標準パターンを用意しておくことにより，まったく同様の方法で認識できる。

(5) 認識結果のうち，子音部分を類似度の上位からいくつか採用する。すでに得られている母音認識結果と合わせて音韻認識結果とする。これらの結果は，セグメントと音韻の両方のあいまいさを持っているため，音韻系列を変形した音韻ラティスの形で表現して出力する。図7.6と同じ例について，得られた音韻ラティスを図7.10に示す。



（＊は音韻が存在していないことを示す。）

図7.10　音　韻　ラ　テ　ィ　ス

7.4　言語処理部の概要[127][131]

言語処理部では，種々の言語情報を利用することによって，音響処理部から受け取った音韻ラティスにおける誤りを訂正して，それを単語系列に変換し，入力の予約内容を決定する。言語処理部は，次の方針に基づいて構成する。

(1)　top－down的な構成を基本とする。言語処理部では，離散的な記号系列を扱い，しかも上位の単語，構文といった概念が簡潔に表現でき，処理の順序関係も明確であるため，top－down的な処理が適していると考えられる。

(2)　言語処理部は，単語認識，構文解析，意味解析の3つの部分からなり，これらのうち，構文解析が処理の中心的な役割をはたす。

(3)　言語情報として，単語認識では音韻変形規則と単語辞書，構文解析では構文情報とプラグマティクス，意味解析では意味情報を利用する。

このような方針に基づく言語処理部の構成を図7.11に示す。図7.11に示す処理の流れの概略は以下の通りである。

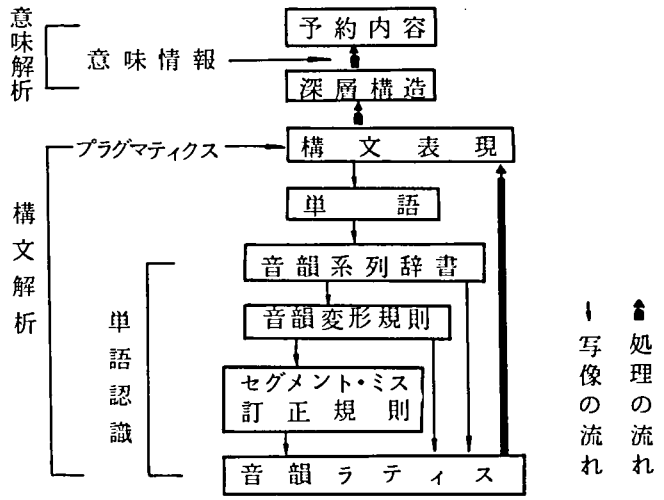


図7.11 言語処理部の構成

(1) まず，プラグマティクによって入力の文節が予測される。

(2) 予測された文節の構文表現の構文解析を実行する。

(3) 構文解析の結果，入力の単語が予測される。

(4) 予測された単語について，音韻変形規則を利用して，音韻ラティスと音韻系列辞書との関連づけによる単語認識を行う。

(5) 単語認識の結果は，次の単語を予測するために構文解析で用いられる。

(6) 構文解析された構文表現（単語系列）から，予約項目のキーワードを抜き出す。

(7) 最後に，意味情報として時刻表を利用して，得られた予約項目の間で推論を行い，最約的な予約内容を決定する。

言語処理部で用いる情報の例として，音韻変形規則を表7.6に示す。また，発駅の構文を表7.7に示す。

表7.6　音韻変形規則

変形の内容	変形規則	個数
音韻の類似性	$x\ v \rightarrow y\ v$	74
調音結合	$u\ x\ v \rightarrow u\ y\ v$	25
音韻の挿入	$u\ v' \rightarrow u\ y\ v'$	17
音韻の削除	$u\ x\ v' \rightarrow u\ v'$	29
母音の無声化	$x \rightarrow y_1\ y_2\ y_3$ [1)]	1
誤セグメントの訂正	$x_1 x_2 x_3 \rightarrow y$ [2)]	8
文頭の子音の挿入	$* \rightarrow *\ y$	5
音節単位の置換	$x_1 x_2 \rightarrow y_1\ y_2$	6

x, x_i ： 変形の対象となる音韻ラティスの音韻

y, y_i ： 音韻系列辞書の音韻

u ： 〃 のyの直前の音韻

v ： 〃 のyの直後の音韻

v' ： 〃 と音韻ラティスの両方に存在する音韻

注1)

x →（無声子音）〔K S T H P〕〔I U〕〔K S T H P〕

注2)

〔* R B M〕〔E I N〕〔* R B M〕→ Y

表7.7　発駅の構文

〔 〕はその中のいずれか一つが選ばれる。
( )はその中のものがあってもなくてもよい。

〔TOKYO
SHINYOKOHAMA
ODAWARA
ATAMI
MISHIMA
SHIZUOKA
HAMAMATSU
TOYOHASHI
NAGOYA
GIFUHASHIMA
MAIBARA
KYOTO
SHINOSAKA〕

(EKI) ((〔KARA
YORI
HATSU〕)) ((〔NO
O
NI
WA
DE
GA
NOO
NONI〕))

7.5 性能評価

システムの性能評価を行うため，男性1名が6項目全部の予約が完了するまで質問回答をくり返す実験を，42種類の予約について行った。また，各予約の最初の会話文章について，音響処理結果，言語処理結果を詳しく検討した。なお，42種類の予約文章を付録2に示す。また，その一部についてソナグラムを付録3に示す。

7.5.1 音響処理結果 (128)

音響処理で用いたVCV標準パターンの種類を表7.8に示す。用いた標準パターンは計468種類である。なお，CV型の標準パターンは，VCV型の標準パターンを時系列の中央で切り，後半分を用いることによって代用する。

表7.8 標準パターンの種類

母音			6個
VCV音節	子音別の分類	VV型	36個
		/w/	6個
		/y, d, zy, ky, hy/	各18個
		/m, n, b, g, r, z, p, t, k, s, h/	各30個
		計	462個

音韻境界抽出結果を表7.9に示す。抽出率は93.4％である。文頭の子音が文中に比較して抽出が困難であることがわかる。挿入誤りには特定の傾向は見られない。脱落誤りのほとんどは，／N／を含めて母音間の境界が検出できなかったものである。

表 7.9　音韻境界抽出結果

	correct	omission error	insertion error
文　頭	86.8％	10.3％	2.9％
文　中	94.4％	3.3％	2.3％
平　均	93.4％	4.2％	2.4％

VCV 音節抽出結果を表 7.10 に示す。表中，“unique”は抽出が一意に定まったもの，“not unique”は一意に定まらなかったものである。“not uniqne”が比較的多いが，これらの多くは拗音，半母音を含んだVCV 音節の抽出が一意に定まらなかったものである。

表 7.10　VCV 音節抽出結果

correct		error
unique	not unique	
51.6％	35.6％	12.8％
87.2％		

音韻認識結果を表 7.11 に示す。表 7.11 には音響処理結果を音韻系列で表現した場合と，音韻ラティスで表現した場合についての比較が示してある。また，音韻認識結果は，結果の中に正しい音韻が含まれている割合と，音韻認識率とで示してある。ここでいう音韻認識率とは，音韻ラティスにおけるように，複数個の音韻の候補で表示された結果の相互情報量を計算し，その情報量から認識率を推定するモデル[148]に基づいて求めたものである。

音韻ラティスによる表現は，通常の音韻系列による表現に比べて，認識率に換算して約 5％多くの情報を含むことがわかり，音響処理結果を音韻ラティスで表現する方が有利と判断される。

表 7.11　音 韻 認 識 結 果

		正しい音韻の含まれる割合		情報量より求めた値	
		音 韻 系 列	音韻ラティス	音 韻 系 列	音韻ラティス
認識率(%)	母　音	85.7	93.4	78	79.5
	子　音	41.6	80.8	37	45
	全　体	63.0	86.9	57.5	62.1

7.5.2　言語処理結果 (127)(131)

予約項目別の単語認識結果と文節認識結果をそれぞれ，表 7.12，表 7.13 に示す。また，表

表 7.12　予約項目別の単語認識結果

	発駅	着駅	発時刻	列車名	等	枚数	動詞	全　体	
正	78	74	169	143	42	20	31	557	76%
(正)	2	2	5	1	2	15	0	27	4%
拒　絶	2	20	42	14	6	8	15	107	14%
誤	2	2	11	4	16	5	2	42	6%
合　計	84	98	227	162	66	48	48	733	100%

表 7.13　予約項目別文節認識結果

	発駅	着駅	発時刻	列車名	等	枚数	全　体	
正	37	33	19	29	27	11	156	78%
(正)	2	0	7	1	1	7	18	9%
拒　絶	1	7	5	2	2	5	22	11%
誤	0	1	2	0	0	1	4	2%
合　計	40	41	33	32	30	24	200	100%

7.12における単語の認識誤りを表7.14に示す。同様に，文節の認識誤りは，予約完了までに質問回答をくり返す過程で生じたものも含めて表7.15に示す。文節認識率は，単語認識率とほぼ

表7.14 単語の誤りの例

（助　詞）	（そ　の　他）
を　→の	9　→分
で　→に	2　→4
の　→のを	10　→時
の　→は	10分　→19分
いき→ゆき	7枚　→米原へ
から→が	4枚　→14時
から→からは	発で　→8分
の　→号	指定　→指定に
を　→のを	取ります→ひかりは

表7.15 意味の誤りの例

（発　時　刻）	
10時25分発　→　9時25分発	
12時19分の　→14時　9分の	
13時30分の　→　7時33分の	
13時53分発の→　7時33分発の	
（枚　数）	
1枚→$\frac{2}{8}$枚	4枚→$\frac{5}{9}$枚
2枚→ひかり	4枚→14時
3枚→5枚	7枚→米原へ
3枚→7時	

等しく78％である。これは，処理結果に拒絶が多く誤りが少ないという系の特性と，単語の誤りでは意味に影響を及ぼさない助詞の誤りが多いことによる。文節の誤りは，すべて発時刻と枚数の文節で生じている。これは，発時刻の文節では多くの単語を含み，かつ，数字のように短い単語をキーワードとしているために元来認識しにくい文節であること，枚数の文節では音響処理部において鼻音のセグメント化と認識がよくないことによる。

言語処理部における誤りを分析すると，音響処理部の誤りで，言語処理部に影響を与えるものは主として，鼻音，拗音，半母音，連続母音，無声化母音の処理であることがわかる。

7.5.3 質問回答結果[130]

表7.16は，42種類の各予約が完了するまでの発声回数を予約項目別に集計した結果である。この表で，発時刻と列車名の中には，実際には発声されないで，推論によって予約されたものも含まれている。また，カッコ内の数字は，その前の回で正しい結果が含まれていたが一意に決らなかったため，“あなたの予約は東京駅からですか？”のようなYES－NO型の音声応答がなされて予約が完了した個数を示す。表7.16より，6項目の予約のうち最初の発声で平均4.6項目が正しく予約されていることがわかる。また，最約に予約できなかった項目については，

平均1.6回の言い直しで予約が完了することがわかる。

表7.16 予約完了までの質問回答の実験結果

		発駅	着駅	発時刻	列車名	等	枚数	全体	
予約完了までの発声回数	1	37	33	28	36	39(8)	21	194	77.0%
	2	3(2)	5	9(5)	5(2)	3(3)	11(6)	36	14.2%
	3	1	1	2			7(5)	11	4.4%
	4	1	3(1)	2	1		1	8	3.2%
	5			1				1	0.4%
	6						2(1)	2	0.8%
合計		42	42	42	42	42	42	252	100%
1回で完了の割合		88%	79%	67%	86%	93%	50%		

( )はYES－NO型質問回答の個数。発時刻，列車名は推論によるものも含む。

7.6 あとがき

会話音声の認識を目ざして作成した会話音声認識システムについて述べた。本システムの認識対象は列車の座席予約に関する会話音声である。システムは音響処理部，言語処理部，音声応答部より構成されている。音響処理部を中心に，各部の構成について述べた。システムの性能評価を行うために，1名の男性発声者が質問回答形式で列車の座席予約をする実験結果を行った。音響処理結果，言語処理結果，質問回答の経過を分析することによりシステムの性能評価を行うと共に問題点を明らかにした。この結果に基づいて各部に改良を加えた会話音声認識の第2次システムについては次章で述べる。

第8章　日本語会話音声認識システムの検討（第2次システム）

8.1　はしがき

第7章では日本語会話音声の認識を目ざして作成した第1次システムの内容および性能評価実験について述べた。本章では第1次システムを改良した第2次システムについて述べる。第2次システムは第1次システムに比較し，処理の精密化，処理の高速化をはかっておりオンラインで動作可能である。本章では音響処理の内容に重点をおいて，システム全体の構成，各部の構成，性能評価実験について述べることとする。

会話音声認識における音響処理に要求される条件として次の4つがあげられる。

(1)　音声のスペクトル分析が精密であること。

(2)　システム構成が柔軟であること。

(3)　発声者に対する適応性がよいこと。

(4)　処理が高速であること。

これら4つの条件を次のような形で実現した。

まず，音声スペクトル分析は，精度の高い分析法である最尤スペクトル分析[16]を採用した。この分析法を用いた音声認識では，第2章で詳しく述べたように，入力音声のスペクトル情報は波形の自己相関関数で抽出し，一方，標準パターンは最尤スペクトルパラメータで蓄えておけばよく，それらのパラメータを用いてスペクトル間の類似度が簡単な計算で求められる利点がある。

次に，システム構成に関しては，従来の音響処理では，特徴抽出，音韻認識を共にbottom－up的手法で行っていたのに対し，音韻認識をtop－down的な手法で行うこととした。これは，音声信号を構成する音韻に関する音声学的知識が比較的明確であるため，これを手がかりにしてtop－down的処理を行うことにより，個々の音韻の性能を考慮した柔軟な処理が行えるためである。

発声者に対する適応性については，会話音声認識が現実のものとなるためには，システムが学習機能を持ち，新しい発声者にすばやく適応できることが必要である。そこで，本システムでは一つの方法として，必要な標準パターンのうち比較的学習の容易な母音標準パターンのみを発声者ごとに作成する方針をとった。

最後にシステムの高速化については，処理の一部を音声情報処理用の高速プロセッサを用いることにより，処理の高速化を目ざした。処理の高速化をはかって作成したオンラインシステムのハードウエア構成については第11章で述べることとする。

8.2 第2次システムの概要[135][136][137]

会話の対象は第1次システムと同じく新幹線の座席予約サービスであり，質問回答をくり返すことにより発声者が意図する予約内容を認識する形式になっている。ただし，対象の範囲は1次システムに比較して拡大している。認識対象の概要を表8.1に示す。予約項目は，日付，発駅，着駅，発時刻，列車名，等，枚数の7項目，駅名の数は28駅，列車本数は181本である。

表8.1 認識対象

タスク	：	列車の座席予約（新幹線）
単語数	：	112個（駅名28個，列車本数181本）
予約項目	：	日付，発駅，着駅，発時刻，列車名，等，枚数
発声の制限	：	文節ごとに0.5秒以上のポーズ
発声例	：	明日の，ひかり191号で，東京から，博多まで， 7時48分発の，指定席を，4枚，予約します。

認識の対象となる112語を表8.2に示す。入力音声は，予約項目の文節ごとに区切って発声するという条件が付加されている。入力音声に含まれる音韻を表8.3に示す。母音は5種類，子音は21種類で，計26種類である。ただし，／*／は連続した母音の境界に挿入する仮想的な子音である。また，撥音／N／は本システムでは母音と同様に扱う。

表8.2 認識の対象となる単語

日付 16	TSUITACHI FUTSUKA MIKKA YOKKA ITSUKA MUIKA NANOKA YOKA KOKONOKA TOKA HATSUKA KYO ASU ASHITA ASATTE HONJITSU
数字 21	12345678910 YO SHICHI KU I RO HA ZYU ZI 100 0 RE
駅名 28	TOKYO SHINYOKOHAMA ODAWARA ATAMI MISHIMA SHIZUOKA HAMAMATSU TOYOHASHI NAGOYA GIFUHASHIMA MAIBARA KYOTO SHINOSAKA SHINKOBE NISHIAKASHI HIMEJI AIOI OKAYAMA SHINKURASHIKI FUKUYAMA MIHARA HIROSHIMA SHINIWAKUNI TOKUYAMA OGORI SHINSHIMONOSEKI KOKURA HAKATA
その他の名詞 15	EKI NICHI JI FUN PUN PPUN HIKARI KODAMA GO GREEN FUTSU SHITEI KEN SEKI MAI
助詞 15	KARA HATSU YORI MADE YUKI IKI E NO O DE NOO NODE WA GA NOWA
動詞 12	YOYAKU ONEGAI ITA SHI MASU ARI MASEN KA DESU DES MAS MOSHIKOMI
その他 5	HAI IIE CHIGAI SO (PAUSE)

計 112個 (音便変化も含む)

表8.3 音声中に含まれる音韻

母音	a, i, u, e, o,
子音	N, w, y, m, n, b, d, g, r, z, zy, p, pp, t, k, kk, ky, s, h, hy, *

会話音声認識システム全体の構成を図8.1に示す。システムは，音響処理部，言語処理部，音声応答部の3つの部分から構成されている。音響処理部と言語処理部で入力音声の認識を行い，認識された予約内容を音声応答部に渡す。音声応答部では，予約の問い合わせや確認を行い，それを合成音声で発声者に伝える。このようにして，システムと発声者の間で質問回答をくり返すことによって正しい予約内容がシステムに認識される。表8.4に質問回答の例を示す。

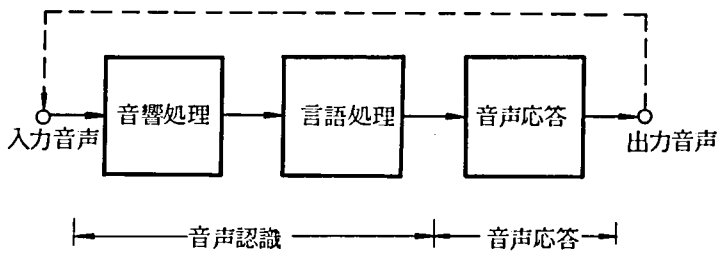


図 8.1　会話音声の認識系の構成

表 8.4 列車の座席予約に関する質問回答の列

C：計算機　S：利用者

C：こちらは新幹線の座席予約センタです。御希望をおっしゃって下さい。
S：東京から，新大阪までの，グリーン券を，お願いします。
C：新大阪までですか？
S：はい，そうです。
C：何日で，何時何分発ですか？
S：1 日で，11 時 48 分発のを，1 枚，下さい。
C：2 枚ですか？
S：いいえ，違います。
C：あなたの予約は，1 日のひかり 109 号で，東京 11 時 48 分発，新大阪までのグリーン券を 1 枚ですね？
S：はい。
C：御希望の指定券はとれました。予約番号は 114 番です。もよりの緑の窓口でお受取り下さい。

8.3 音響処理 (92)(132)(133)(134)

8.3.1 音響処理の構成

音響処理の構成を図 8.2 に示す。入力音声に対して，特徴抽出および音韻認識が行われ，処理結果は音韻ラティスの形式で表現される。音韻ラティスは図 8.3 に示すように，認識処理におけるセグメント化のあいまいさと音韻認識のあいまいさを含んだ表現形式である。著者は，

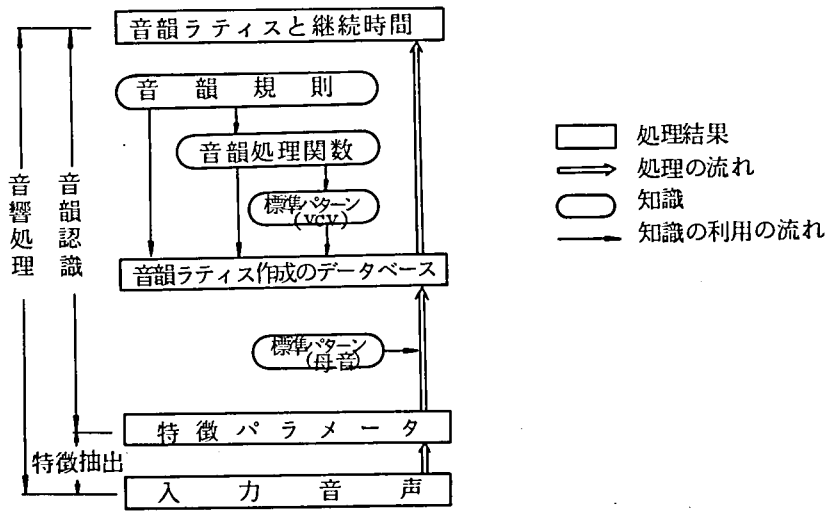


図 8.2　音響処理の構成

第 4 章で述べた単語音声認識や，第 5 章で述べた連続単語音声認識において，セグメント化やセグメントの認識の段階で生じる誤りが認識結果に決定的な影響を与えるのをさけるため，処理結果にあいまいさを許した表現形式をとった。音韻ラティスは，この考え方を発展させたものである。音韻ラティスという表現形式をとることにより，認識結果を 1 時点について 1 種類の音韻記号のみで表現するいわゆる音韻系列よりも，柔軟性に富み，前章でも示したように多くの情報を持った表現が可能になる。したがって，音韻ラティスは音韻系列よりも有効な表現

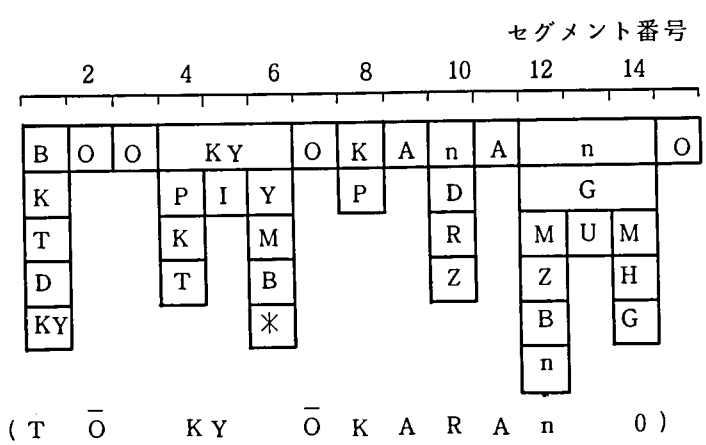


図 8.3 音 韻 ラ テ ィ ス

形式である。他の研究機関，たとえばBBNにおける音声理解系の研究においても処理結果を同じような形式で表現している。[105] 音響処理部から言語処理部へ送られる情報は，音韻ラティスと各音韻の継続時間である。

音響処理では，入力音声の特徴抽出を行って，音韻ラティス作成用のデータベースを作るまでの処理はbottom－up的に行われる。これは，この部分の処理が音声波形から直接導かれる低次のパラメータを扱う処理であるため，bottom－up的に一様に処理する方が能率的であると考えられるからである。これに対し，データベースから最終的な音韻ラティスを作成する処理はtop－down的に行われる。top－down的処理を導入したのは，従来の音響処理における特徴抽出→セグメント化→音韻認識というbottom－up的処理のみでは次のような欠点があるためである。

(1) 会話音声中には，継続時間，音韻から音韻への変化速度等が異なる種々の音韻が含まれている。これらの音韻に対し一様な処理を行うと，拗音，鼻音，連続母音等の音韻の処理が困難である。

(2) 処理の方向が，セグメント化→音韻認識という一方向的なものであると，セグメント化の段階で生じた誤りを後で訂正することが困難である。

したがって，音響処理においてもbottom－up的処理のみでは不十分であるといえる。従来，言語処理においてはtop－down的処理が主としてとられてきたのに対し，音響処理においてはbottom－up的処理を行う場合がほとんどであった。これは言語処理においては構文などの上位概念が明確であったのに対し，音響処理では，その上位概念である音韻レベルでの規則が明確ではなく，表現が困難であるという理由による。しかしながら，音韻レベルでの知識としては，従来から音声学的な知識が種々得られており，これは比較的容易に表現できる。また，多量のデータを使って音響処理実験を行う過程で得られる知識は音韻レベルのものが多い。そこで，ここでの音響処理では，音響レベルの上位概念である音韻に関する知識をtop－down的に利用して音韻認識を行うこととした。このようなシステム構成をとると次の利点があると考えられる。

(1) 音声学的な知識や，その他われわれが持っている音韻レベルでの知識をシステムの中へ組み込むことが容易になり，システム構成が柔軟になる。

(2) 音響処理と言語処理を単にserialに結合するのでなく，将来，より密な結合が可能になる。

以下，8.3.2では特徴抽出からデータベース作成までの処理の内容について述べる。8.3.3ではtop－down的な音韻認識の手法について述べる。

8.3.2 特徴抽出とデータベース作成

図8.4に，入力音声を特徴パラメータに変換し，音韻ラティス作成のためのデータベースを作り上げる過程を示す。

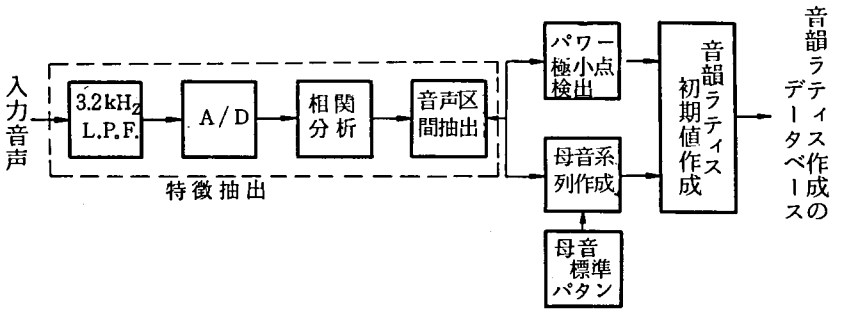


図8.4　特徴抽出とデータベース作成の手順

8.3.2.1 特徴抽出

入力音声は3.2kHz低域通過フィルタに通した後，標本化周波数8kHzのAD変換器で12ビットのディジタル音声に変換される。ディジタル音声は15 msecのフレームに分割され，各フレームごとに音声の特徴量として，波形の自己相関関数

$$\boldsymbol{v}=(v_0,\ v_1,\ \cdots\cdots\cdots\cdots,\ v_p) \qquad (8.1)$$

を求める。なお，式（8.1）において，v_0は音声パワーを表わす。音声区間抽出のためのパラメータとしては，音声パワーv_0と第1次の自己相関係数v_1/v_0を用いる。これらのうち，第1次の自己相関係数はスペクトル全体の傾きを示すパラメータであり，文頭の無声子音を検出するための情報として用いる。音声区間の抽出は次の手順で行う。

音声パワーに関する2種類の閾値P_1，P_2（$P_1 > P_2$）およびv_1/v_0に関する閾値P_3を定める。P_1より大きい音声パワーが検出されると，その10フレーム前から再び分析し，

$$v_0 > P_2 \text{ かつ } v_1/v_0 < P_3 \qquad (8.2)$$

であるフレームを音声区間の始端とする。また，P_1以下のパワーが30フレーム（450 msec）以上続いたとき，最初に閾値以下になったフレームを音声区間の終端とする。

8.3.2.2 母音系列作成

15 msecのフレームごとに母音標準パターンとのマッチングを行い，入力を母音の記号列に変換する。母音標準パターンは最尤スペクトルパラメータの形で蓄えておく。母音/x/の標準パターンを

$$(A_{x0},\ A_{x1},\ \cdots\cdots,\ A_{xp}) \qquad (8.3)$$

とする。また，入力音声の第iフレームの自己相関関数を

$$(v_{i0},\ v_{i1},\ \cdots\cdots,\ v_{ip}) \qquad (8.4)$$

とする。式（8.3）と式（8.4）の類似度として第2章で定義した類似度を用いる。これは次式で与えられる。

$$l(i,\ x) = -\log\left\{ \sum_{\tau=0}^{p} v_{i\tau} A_{x\tau} \right\} \qquad (8.5)$$

式（8.5）を最大にする/x/，すなわち類似度第1位の母音を用いて母音系列を作成する。同様にして，類似度第2位の母音を用いた母音系列も作成する。

8.3.2.3 パワー極小点検出

子音の多くは，前後の母音に比較して，口がせばめられる，共鳴が小さい等の理由で音声パワーが小さくなる。この性質を利用して子音部分に対応するパワーの極小点を検出する。図8.5に処理の例を示す。2種類の閾P_T, Q_Tを定める。まずパワーがP_Tより小さくなる区間を検出し，区間の各フレームに記号Cを対応づける。次に，残された区間においてパワーの極小点を検出する。そして，前後の極大点とのパワーの差が閾値Q_Tより大きい極小点を与えるフレームを検出し，そのフレームに記号Dを対応づける。

8.3.2.4 音韻ラティス初期値作成

母音系列，およびパワーの極小点を使って音韻ラティスの初期値を作成する。これは次の手順で行う。

(1) パワーの極小点記号Cの区間を無音区間として，同じ記号Cであらわす。

(2) 残された区間において，類似度第1位の母音系列をしらべ，同じ母音が4個以上連続してあらわれ，かつ，パワーの極小点記号Dが途中にない場合は，その区間に母音が存在すると

して，記号Vであらわす。

(3) 残された区間は過渡区間として，記号ーであらわす。

(4) 母音区間においては，類似度第1位の母音を母音認識結果とする。

以上の手順で音韻ラティスの初期値が作成される。この例を図8.5に示す。図からわかるように，この段階での音韻ラティスは，子音認識が行われておらず，かつ，子音を母音と判定するなどの誤りが存在している。したがって，以上の手順で得られた，自己相関関数の時系列，音声パワーの極小点，母音系列，および音韻ラティスの初期値は，音韻ラティス作成のデータベースとして音韻認識部へ送られ，より精密な処理が行われる。

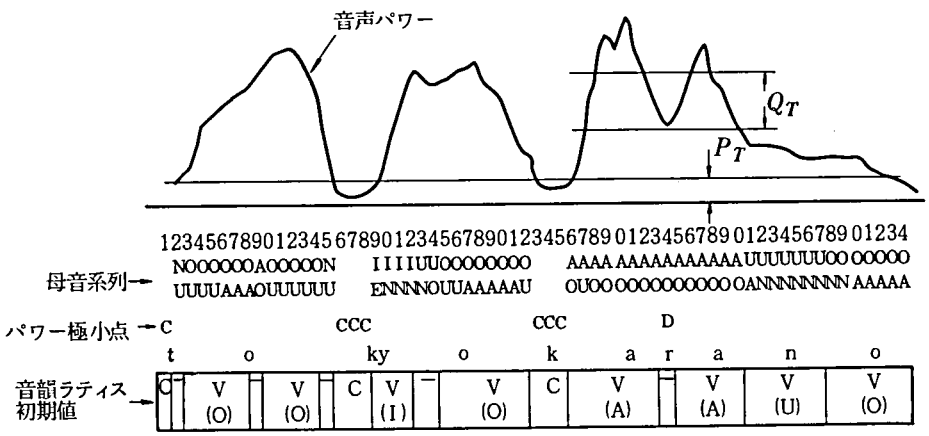


図8.5 音韻ラティス初期値作成の例

8.3.3 音韻認識

音韻認識は，次に示す3種類の知識をtop－down的に利用することにより行う。

(1) 音韻規則

(2) 音韻処理関数

(3) VCV音節標準パターン

これらの知識の内容，および，それを用いた処理の手順について説明する。

8.3.3.1 音韻規則

連続音声中では，音韻が前後の音韻の影響を受けてその音響的性質が変化する，いわゆる，調音結合の現象が生じる。そのため，音韻ラティスの初期値においては，母音の脱落，子音を

母音とする誤り，母音の認識誤り，などが存在している。調音結合は，調音器官が時間の関数として連続的に動作することにより生じるものであり，発声者に共通した規則性が存在するものと思われる。したがって，音韻ラティス上の誤りの傾向にも規則性が存在すると考えられる。音韻規則は，このような誤りの規則性を見出し，それを音韻ラティス上の3つの記号の系列として表現したものである。規則を3個の記号の系列で表現したのは，調音結合の及ぶ範囲が，音韻とその前後の音韻を含めた3音韻程度であることが知られているためである。[149] 音韻規則は，表8.5にある48種類が用意してある。たとえば，鼻子音は母音/u/，または発音/N/に誤ることが多いため，これをVUV，VNVという規則で表現している。

表8.5　音　韻　規　則

番号	規　則	インデックス	対応する処理関数	処理内容
1	C — C	0	文頭・文末処理	— →V
	C — *	0		— →V
	C V C	1		V →—
2	V — V	0	過渡部処理	— →V
	— V *	0		— →V
3	V V *	0	母音処理	V →V
4	C — V	0	無声化母音処理	— →V
	C N V	1		N →V
	C U V	1		U →V
	V — C	0		— →V
5	N — *	0	撥音処理	— →V
	U — *	0		— →V
	N V *	1		N →—
	U V *	1		U →—
	N N *	1		N →—
	V U *	1		U →V
	V N *	1		N →V
6	V O A	1	連続母音処理	O →—
	V A O	1		A →—
	V A A	1		A →—
	V E I	1		E →—
	A E —	0		— →V
	A — *	0		— →V
	— E A	0		— →V
	E — *	0		— →V
7	V U V	1	鼻音処理	U →—
	V N V	1		N →—
8	V I A	1	拗音処理	I →—
	V E A	1		E →—
	V I O	1		I →—
	V E O	1		E →—
	V I U	1		I →—
	V E U	1		E →—
	C I A	1		I →—
	C E A	1		E →—
	C I O	1		I →—
	C E O	1		E →—
	C I U	1		I →—
	C E U	1		E →—
	I E N	1		IE→—
	I E U	1		IE→—
	I E V	1		I →—
	E N V	1		E →—
	I N V	1		I →—
	I E C	1		I →—
	I N C	1		I →—
	C I *	1		I →—
	C E *	1		E →—

(1) 記号＊は音韻ラティスのどの記号でもよいことを示す。 (2) インデックスは，規則を適用するときに音韻ラティス上の過渡区間をとばしてもよいか(1)否か(0)を表わす。 (3) 処理内容とは，左側の記号を右側の記号に変えることを意味している。たとえば(C－C)という規則で抽出された子音区間(－)は母音区間(V)にかえられることを意味している。

処理を行うには，まず，音韻規則に合致する部分を音韻ラティス上で探す。規則に合致する部分が見つかると，各規則に対応して，どの音韻処理関数を用いて，どのセグメントをチエックすべきかが決る。

8.3.3.2 音韻処理関数

音韻処理関数は，データベースの情報を用いてセグメントのチエックを実行し，音韻ラティスに変更を加える。音韻ラティスに加える変更としては，次の3種類が可能である。

(1) 新しいセグメントの追加…音韻ラティスに新しいセグメントを追加する機能であり，図8.6に示す3種類のものがある。まず，第1は音韻ラティス上の母音セグメントに子音セグメントを追加するもので，鼻音等が母音として検出された誤りを訂正する際に用いる。第2は子音セグメントに母音セグメントを追加するもので，母音の無声化等により母音が子音として検出された誤りを訂正するために用いる。第3は2つの母音セグメントに対して1つの子音セグメントを追加するもので，拗音において過渡部分を2つの母音として検出した誤りを訂正する際に用いる。

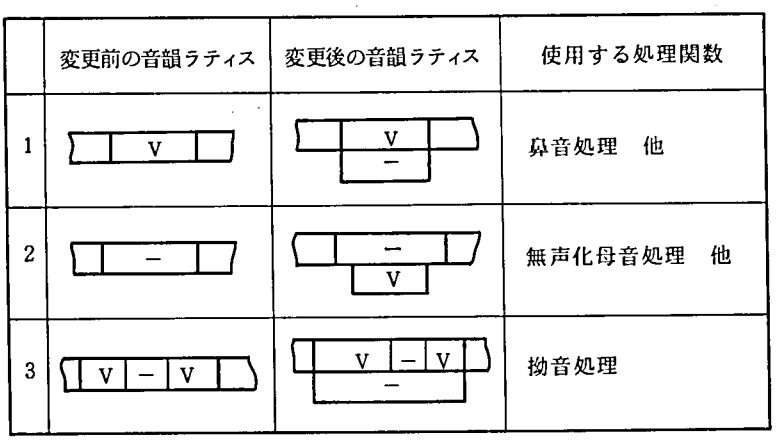


図8.6 セグメントの追加機能

(2) セグメントの削除…音韻ラティスのセグメントを削除する機能である。

(3) セグメントの音韻情報の追加…子音認識を行い，過渡区間，無音区間の音韻を決定したり，母音認識結果が不確実なセグメントについては，母音の候補を追加する等の機能である。

処理関数は，表8.5に示した8種類が用意してある。それぞれの処理関数の内容について簡

単に説明する。

(1) 文頭・文末処理

文頭，文末で無声子音と共に発声される母音は短くなりやすく，過渡部と判定される誤りが多い。また，文頭の有声破裂音はbuzz音を伴うことが多く，このbuzz音を母音と判定する誤りが生じやすい。このような誤りの回復をはかるための処理関数である。

(2) 過渡部処理

発声速度が速い場合は，母音系列が不安定となり，母音区間が長い過渡区間として抽出されることがある。このような場合に，母音系列上で同じ母音が3個連続してあらわれれば母音が存在すると判定することにより，母音抽出誤りの訂正をはかる処理関数である。

(3) 母音処理

長母音が2つの母音にセグメント化されてしまう場合がある。このような誤りを訂正する処理関数である。

(4) 無声化母音処理

無声子音の前後では母音が短くなりやすく，過渡部と判定される誤りが多い。したがって，無声子音の前後では，母音系列上で同じ母音が2個連続してあらわれれば母音が存在すると判定すると共に，抽出された母音が不確実であれば，類似度第2位の母音系列を用いて母音の候補をもう1個追加する。このような方法によって，母音の抽出誤りを少なくするための処理関数である。

(5) 撥音処理

撥音/N/は本システムでは母音と同等に扱っているが，母音/u/との混同が生じやすい。撥音は文頭，および子音の後には発声されないため，この性質を利用して撥音の認識誤りを少なくするための処理関数である。

(6) 連続母音処理

複数個の母音が連続して発声された場合，口の開きの大きい母音は強く発声され，これに対し，/u/，/i/のように口の開きの小さい母音は弱く発声される。したがって，たとえば，/ai/と発声すると音韻ラティス上ではAEという系列があらわれ，Iが検出されないことが多い。また，半母音/y/，/w/は，/i/+母音，/u/+母音という関係にある連続母音と考えられ，同じように取扱える。このような場合の処理を行う関数である。

(7) 鼻音処理

鼻音は定常的で，かつ，継続時間の長い子音であるため，母音に誤ることが多い。この誤り

を訂正するための処理関数である。たとえば，音韻規則VUVに合致する部分が音韻ラティス上に見つかると，鼻音処理関数は，Uに対応するセグメントにおいて，継続時間のチエック，VCV音節標準パターンとのマッチングの度合のチエック等を行い，そのセグメンが鼻子音かどうか判断し，結果に応じて音韻ラティスに変更を加える。

(8) 拗音処理

/zy/, /ky/, /hy/等の子音はゆるやかに変化しながら長く継続する子音であり，他の子音と性質が異なっているために，従来の一様なbottom−up的処理方法では検出が困難であった。しかしながら，音韻ラティス上で見るとこれらの子音は特徴的なパターンを示す。たとえば，/zyu/は音韻ラティス上では，IU，IEU等の系列としてあらわれる。したがって，拗音処理関数は，音韻規則により検出された拗音の候補点において，VCV音節標準パターンとのマッチングを行い，拗音が存在するかどうかをたしかめるという手法で拗音処理を行う。

8.3.3.3 VCV音節標準パターン

VCV音節標準パターンは，音韻処理関数の中でVCV音節単位で子音認識を行うときに用いる。子音の認識単位としてVCV音節を用いるのは，子音は前後の母音の影響を受けやすい不安定な音韻なので，前後の母音と組にしたVCV音節を単位として認識を行った方が誤りが少ないと考えられるためである。著者はこのような考え方に立って，VCV音節を単位とした単語認識，連続単語認識を行って，その有効性をたしかめた(第4章，第5章)。したがって，本システムにおいても，VCV音節を単位とした子音認識法を用いる。VCV音節標準パターンと入力のマッチングの方法を次に示す。

入力のセグメントが相関関数の時系列

$$(\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots\cdots,\ \boldsymbol{v}_N) \tag{8.6}$$

$$\text{ただし } \boldsymbol{v}_i = (v_{i0},\ v_{i1},\ \cdots\cdots,\ v_{ip}) \tag{8.7}$$

として与えられているとする。$\boldsymbol{v}_i$より求めた最尤スペクトルパラメータを

$$\boldsymbol{B}_i = (B_{i0},\ B_{i1},\ \cdots\cdots,\ B_{ip}) \tag{8.8}$$

とする。これに対し，VCV音節標準パターンは最尤スペクトルパラメータの時系列

$$(\boldsymbol{A}_1,\ \boldsymbol{A}_2,\ \cdots\cdots,\ \boldsymbol{A}_M) \tag{8.9}$$

$$\text{ただし}\quad \boldsymbol{A}j = (A_{j0},\ A_{j1},\ \cdots\cdots,\ A_{jp}) \tag{8.10}$$

としてたくわえられている。入力の第 i フレームと標準パターンの第 j フレームの類似度として，第2章で述べた類似度

$$l(i, j) = -\log\left\{\sum_{\tau=0}^{p} v_{i\tau} A_{j\tau}\right\} + \log\left\{\sum_{\tau=0}^{p} v_{i\tau} B_{i\tau}\right\} \qquad (8.11)$$

を用いる。式（8.11）の右辺第2項は標準パターンに関係しない項であり，これを省略すると

$$l(i, j) = -\log\left\{\sum_{\tau=0}^{p} v_{i\tau} A_{j\tau}\right\} \qquad (8.12)$$

となる。

拗音の標準パターンとのマッチングのように，マッチングの絶対的な度合を求める必要がある場合には式（8.11）の類似度を用いる。これに対し，いくつかの標準パターンの中から入力に最も良く似た標準パターンを求める場合は，標準パターンに無関係な項を除いた相対的な値で良いため，式（8.12）の類似度を用いる。この類似度を i と j のすべての組合せについて計算し，類似度行列

$$LM = \left\{l(i, j)\right\} \qquad (8.13)$$

$$(i = 1, 2, \cdots\cdots, N\ ;\quad j = 1, 2, \cdots\cdots, M)$$

を作成する。入力と標準パターンの類似度は，入力の各フレームを標準パターンのいずれかのフレームに対応づけた場合の，各フレームごとの類似度の和として次式で定義する。

$$\max_{f}\left\{\sum_{i=1}^{N} l(i, f(i))\right\} \qquad (8.14)$$

式（8.14）の f は入力と標準パターンの各フレームの対応づけを示す関数であり，次の条件を満すものとする。

$$\text{(i)}\quad 1 \leqq f(1) \leqq f(N) \leqq M \qquad (8.15)$$

$$\text{(ii)}\quad f(i) - f(i-1) = 0,\ 1,\ 2$$

式（8.15）の(i)は，入力を標準パターンの一部に対応づけることを示している。(ii)は，入力の各フレームを2フレーム以内の伸縮をゆるして標準パターンのフレームに対応づけることを示している。式（8.14）は，式（8.13）の類似度行列上で第1例から第 N 列へ至る類似度和最大のパスを求める問題であり，DPを用いて容易に求めることができる。この様子を図8.7に示す。VCV音節標準パターンは単独に発声したVCV音節から作成するため母音定常部を含んでいる。

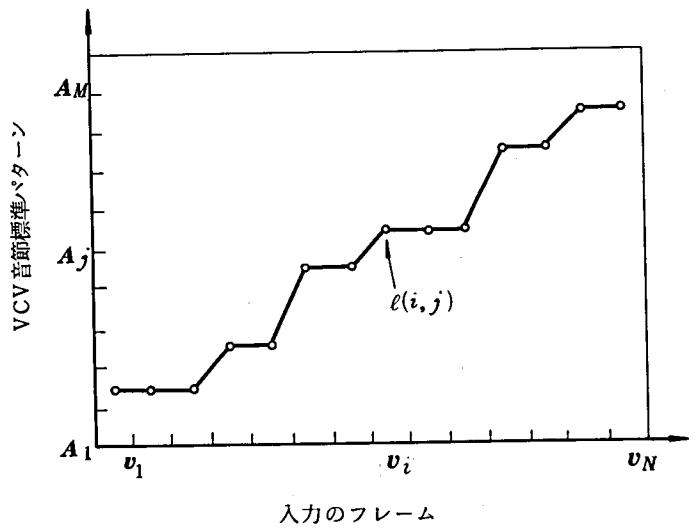


図 8.7 端点フリーDPマッチング法によるVCV音節の認識

これに対して，連続音声中から切り出したVCV音節は母音定常部を含んでいないことが多い。したがって，この両者をマッチングするには，ここに示したように入力を標準パターンの一部に対応づける方法が有効である。この方法を端点フリーDPマッチング法と呼ぶ。端点フリーDPマッチングについての詳しい説明は第6章で述べたので，ここでは省略する。

8.3.3.4 音韻認識の手順

以上述べた3種類の知識を用いて音韻認識を行う手順は次の通りである。

(1) 表8.5に示した音韻規則を順に音韻ラティスに適用する。音韻ラティス上で規則と合致する部分が見つかると，その位置およびチエックすべき内容を引数として，対応する音韻処理関数が呼ばれる。この様子を図8.8に示す。

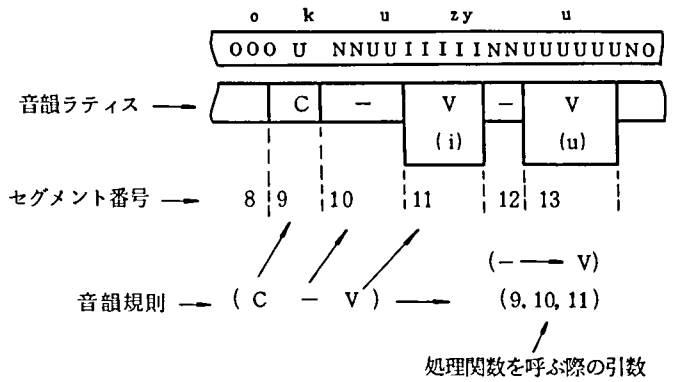


図 8.8　音韻規則と音韻ラティスのマッチング

(2)　呼ばれた音韻処理関数は指示されたセグメントのチエックを行い，その結果に応じて，音韻ラティスに変更を加える。

(3)　以上の手順をくり返すことにより処理が行われる。音韻ラティスに適用する音韻規則の順序は次の通りである。

1 → 2 → 3 → 4 → 5 → 6 → 7 → 8
↑_____________________|

（番号は表 8.5 の規則の番号に対応している）

この順序は，できるだけ少ない処理のくり返しで正しい音韻ラティスが作られるように，実験により定めたものである。本システムでは，くり返しの回数は 2 回としている。音韻認識過程の例を図 8.9 に示す。

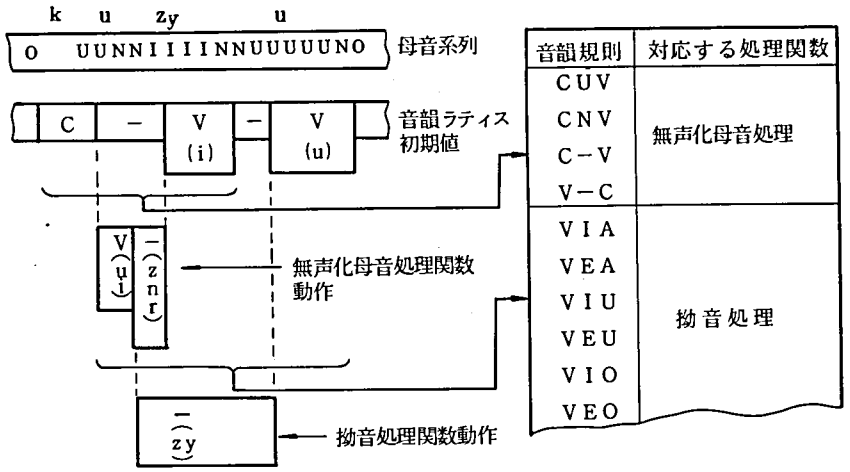


図8.9 音韻認識過程の例

8.3.4 母音の学習方法

不特定多数の発声者に対して認識システムが正しく動作できるためには，システムが学習機能を持つ必要がある。本システムで必要な標準パターンは，母音標準パターンとVCV音節標準パターンの2種類である。このうち比較的学習の容易な母音については，発声者ごとに標準パターンを作成する方法をとっている。また，VCV音節標準パターンは，固定したものを発声者に共通に用いる。母音標準パターンの学習は次の手順で行う。

(1) あらかじめ，特定の発声者の音声を最尤スペクトルパラメータの時系列としてたくわえておく。これを基準サンプルと呼び，式（8.9）で表現される。また，基準サンプルから視察により抽出した母音$/x_i/$の区間を（i_1，i_2）とする。

(2) 新しい発声者は同じ内容の音声を発声する。この音声信号は相関関数の時系列で表現される。これを学習サンプルと呼び式（8.6）であらわす。

(3) 基準サンプルと学習サンプルのDPマッチングを式(8.11)〜式（8.15）の手順で行う。ただし，基準サンプルと学習サンプルは，始点，終点を一致させてDPマッチングを行うので，関数fには式（8.15）の代りに次の条件を課す。

$$\text{(i)}\quad f(1)=1,\ f(N)=M \tag{8.16}$$

$$\text{(ii)}\quad f(i)-f(i-1)=0,\ 1,\ 2$$

(4) (3)で求めた関数fは，基準サンプルと学習サンプルの各区分の対応づけを示す関数である。基準サンプルの母音/x_i/の区間（i_1, i_2）はあらかじめわかっているから，（i_1, i_2）に対応する学習サンプルの区間（j_1, j_2）を求めれば，これが学習サンプルから抽出された母音/x_i/の区間である。

(5) 学習サンプルの区間（j_1, j_2）中の相関関数を平均し，これから得られる最尤スペクトルパラメータを新しい発声者の母音/x_i/の標準パターンとする。

8.4 言語処理[93]

言語処理の概要について述べる。言語処理の構成を図8.10に示す。言語処理は単語認識，構文解析，推論の3つの部分より構成されている。単語認識は構文解析の中に組み込まれている。

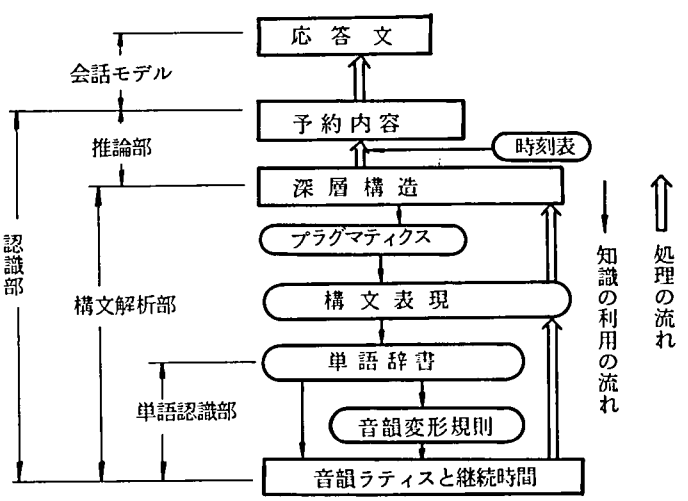


図8.10 言語処理の構成

8.4.1 単語認識

単語認識部では構文解析によって予測された単語と音韻ラティスのマッチングが行われる。マッチングの際，単語辞書と書き換え規則の2種類の言語情報が用いられる。単語辞書中では認識対象の単語が音韻系列の形で蓄えられている。書き換え規則は，音韻ラティス中の誤りを訂正するために用いられる。表8.6に示した14種類の書き換え規則を用意することにより，種

々の調音結合に対応できるようになっている。規則の数は全体で477個である。

表8.6 音韻変形規則

	変形規則の内容	変形規則の型	継続時間
母音に関する規則	短母音から長母音への変形	$V_l \rightarrow V_d$	$t(V_l) \geqq D$
	調音結合による母音の誤りの訂正	$C_d V_l C_{ld} \rightarrow C_d V_d C_{ld}$	$t(V_l) \geqq D$
	母音の挿入(1)	$V_l \rightarrow V_d V_l$	$t(V_l) \geqq D$
	母音の挿入(2)	$V_{ld} \rightarrow V_{ld} V_d$	$t(V_{ld}) \geqq D$
	母音と音節との間の誤りの訂正	$C_l V_l \rightarrow V_d$	
	同じ母音ではさまれた子音の挿入	$V_{ld} \rightarrow V_{ld} C_d V_{ld}$	$t(V_{ld}) \geqq D$
子音に関する規則	子音の類似性(1)*	$C_l V_d \rightarrow C_d V_d$	$t(C_l) \geqq D$
	子音の類似性(2)*		$t(C_l) \leqq D$
	子音の挿入	$V_d V_{ld} \rightarrow V_d C_d V_{ld}$	—
	調音結合による誤りの訂正	$C_l V_l C_l' \rightarrow C_d V_d C_d'$	—
	文頭の子音の挿入	$* \rightarrow * C_d$	—
	拗音のセグメント化誤りの訂正	$C_l V_l C_l' \rightarrow C_d$	$t(V_l) \leqq D$
その他	音韻の続みとばし	$P_d P_l P_{ld} \rightarrow P_d P_{ld}$	$t(P_l) \leqq D$
	母音の無声化の訂正	$C_l \rightarrow C_d V_d C_d$	$t(C_l) \geqq D$

* 子音間の類似性の規則を，継続時間情報の利用方法の違いにより，2種類の規則にわけた。

(1) 添字 l, d, ld は，l：音韻ラティスの音韻，d：単語辞書の音韻，ld：ラティスと辞書の両方に存在する音韻，を表わす。

(2) $t(x)$ は音韻 x の継続時間を表わす。

マッチングは，tree searchの手法を用い，depth－first, left－to－rightで行われる。単語認識に要するステップ数を削減するため，次の手法を用いている。

(1) 書き換え規則が適用される際は，ペナルティが課せられる。ペナルティはマッチング処理の進行に伴って累積され，累積値が閾値をこえると，その処理は終了し，他の候補についてマッチング処理を行う。

(2) tree searchの過程は記憶しておき，同じ処理をくり返さないための情報とし用いる。

(3) 単語境界はあいまいな場合が多い。あいまいな単語境界を別々に扱わずに一括して処理することによりマッチング処理量を削減することができる。

(4) 単語認識の結果を記憶しておく。後の処理で同じ単語が予測された場合は，マッチング処理を行う必要はなく，記憶してある単語認識結果を参照するだけで良い。

単語認識が成功した場合には，次の単語が予測される。もし失敗した場合は別の単語が予測される。このようにして，成功するまで単語の予測が続けられる。

8.4.2 構文解析

構文解析部では，入力音声の文節，単語が予測される。ここでは，構文およびプラグマティクスの2種類の言語情報が用いられる。

構文はリスト構造で表現されている。その例を図8.11に示す。構文のリストをparsingすることにより単語の予測が行われる。

parsingの過程で(X)(ただしXは辞書の表現形式) の形の部分リストが見出されると，単語Xが予測される。

プラグマティクスは発声者とシステムの間でかわされる質問回答に関する会話モデルの知識である。この知識も図8.12に示すようにリストの形で表現されており，入力音声の文節を予測するのに用いられる。プラグマティクスに関するリストのparsingを行い，(# X)という形のリストが発見されると，Xという名のリストのparsingが開始される。

なお，言語処理部で用いる言語情報を付録4に示す。

```
DATE : ((OR (* TSUITACHI) (* FUTSUKA) (* MIKKA) (* YOKKA) (* ITSUKA) (* MUIK
       A) (* NANOKA) (* YOKA) (* KOKONOKA) (* TOKA) (* HATSUKA) (* KYOASU)
       ((OPT (* 2)) (* 10) (OR ((OR (* 1) (* 2) (* 3) (* 5) (* 6) (* SHICHI) (* 7) (* 8)
       (* KU) ((* NICHI)) (* YOKKA))) ((* 3) (* 10) (OPT (* 1)) (* NICHI))) (* JOSHI))
STARTING-STATION : ((*EKIMEI) (OPT (* EKI)) (OR (* KARA) (* HATSU) (* YORI))
       (* JOSHI))
ARRIVING-STATION : ((* EKIMEI) (OPT (* EKI)) (OR (* MADE) (* YUKI) (* IKI) (SEM
       (LATTICE 3 5) (* E))) (* JOSHI))
STARTING-TIME : ((* SUJI6-22) (* JI) (OPT (* PAUSE)) (* SUJI0-59FUN) (OPT (* HAT
       SU)) (* JOSHI))
NAME-OF-TRAIN : ((OR ((* HIKARI) (ORT (* PAUSE)) (* SUJI1-199)) ((* KODAMA)
       (OPT (* PAUSE)) (* SUJI200-299))) (* GO) (* JOSHI))
SEAT-CLASS : ((OR (* SHITEI) (* FUTSU) (* GREEN)) (OPT (OR (* KEN) (* SEKI)))
       (* JOSHI))
NUMBER-OF-TICKETS : ((* SUJI1-9) (* MAI) (* JOSHI))
YES-NO : (OR (* HAI) (* IIE) ((* SO) (* DESS)) ((* CHIGAI) (* MASS)))
VERB : (OR ((OR ((OR (* YOYAKU) (* ONEGAI)) (OPT (* ITA)) (* SHI)) ((* MOSHIKO
       MI) (OPT ((* ITA) (* SHI))))) (* MASS)) ((* ARI) (OR (* MASU) (* MASEN))
       (* KA)) (* DESS))
```

図 8.11　構文表現（リストの名前のうち，JOSHI，EKIMEI，SUJI6−22，SUJI0−59 FUN，SUJI2−5，SUJI1−199，SUJI200−299，KYOASU，MASS，DESS，NICHI は構文のリスト表現を，その他は単語辞書の音韻表記したリスト表現を表わす）

```
(SEM (YES-NO) (# YES-NO) SEM (VERB) SEM (LATTICE 4 20) (# VERB) SEM (SYNTAX
1) SEM (LATTICE 4 24) (# DATE) SEM (SYNTAX 2) SEM (LATTICE 6 26) (# STARTING-
STATION) SEM (SYNTAX 3) SEM (LATTICE 8 26) (# ARRIVING-STATION) SEM (SYNT
AX 4) SEM (LATTICE 12 64) (# STARTING-TIME) SEM (SYNTAX 5) SEM (LATTICE 10 64)
(# NAME-OF-TRAIN) SEM (SYNTAX 6) SEM (LATTICE 2 18) (# SEAT-CLASS) SEM (SYN
TAX 7) SEM (LATTICE 4 14) (# NUMBER-OF-TICKETS))
```

図 8.12　プラグマティクスのリスト表現

8.5 会話モデル [94][136]

座席予約の質問回答を行うための会話モデルについて説明する。

会話状態として，(1)予約開始の案内，(2)まだわからない予約項目の問合せ，(3)あいまいな予約項目の確認，(4)全予約項目の確認，(5)予約項目間の矛盾の訂正，(6)その他の訂正要求，(7)予約終了の案内，の7つの状態を設定した。これらの会話状態の説明，および各状態に対応する応答文の例を表8.7に示す。

表 8.7 会話状態の説明と各状態における応答文の例

会話状態	説明	応答文（下線部は可変）	
予約開始の案内	予約の開始を発声者に告げる。	こちらは新幹線の座席予約センターです。 ご希望をおっしゃって下さい。	固定 (1)
未定の予約項目の問合せ	まだ認識していない予約項目を問合わせる。	何日ですか？ どこからですか？ どこまでですか？ 何時何分発ですか？ どの列車ですか？ 普通席・グリーン席のどちらですか？ 何枚ですか？ 何日でどこからですか？ どこからでどこまでですか？ どこまでで何時何分発ですか？ 何時何分発でどの列車ですか？ どの列車で普通席グリーン席のどちらですか？ 普通券・グリーン席のどちらで何枚ですか？	固定 (7) 可変 (6)
あいまいな予約項目の確認	あいまいな予約内容を確認する。	1日ですか？ 東京駅からですか？ 博多駅までですか？ 9時40分発ですか？ 9時台ですか？	可変 (5)
全予約項目の確認	すべての予約項目が矛盾なく一意に認識されたとき予約内容を確認する。	あなたの予約は3日のひかり191号で 東京駅7時48分発名古屋駅までの 普通券を3枚ですね？	可変 (1)
予約項目間の矛盾の訂正の要求	予約内容間に矛盾があるために，発声者に訂正を要求する。	ご希望の予約はひかり41号東京駅 9時30分発博多駅までと了解しましたが， 該当する列車がございません。当方の聞き違いと思いますので訂正をお願い致します。	可変 (1)
その他の訂正の要求	発声者に訂正を要求する	訂正をお願い致します。	固定 (1)
予約完了の案内	予約が完了したことを発声者に告げる。	ご希望の指定券はとれました。予約番号は4番です。もよりの緑の窓口でお受け取り下さい。	可変 (1)

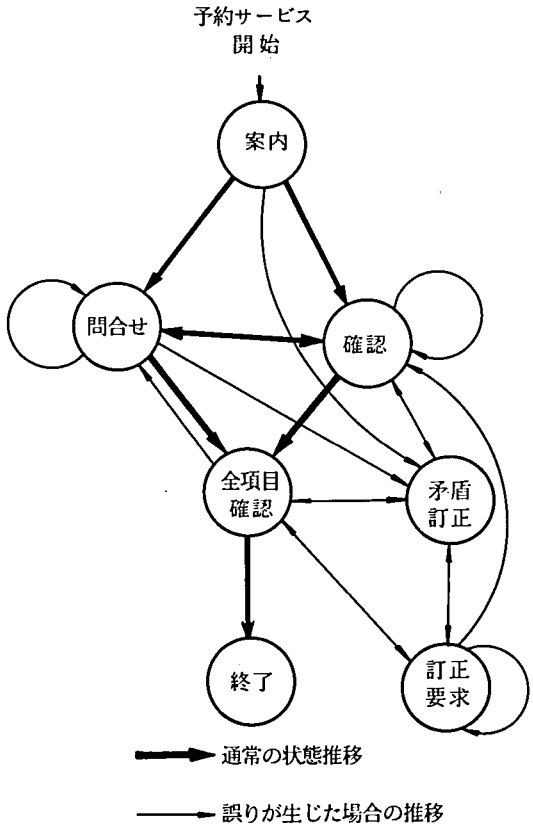


図 8.13　会 話 状 態 推 移 図

会話状態の推移図を図 8.13 に示す。会話状態の推移は

$$\left\{\begin{array}{l}\text{現在の会話状態}\\ \text{予約内容の把握状態}\\ \text{認識結果（はい，いいえ，リジェクト）}\end{array}\right\} \Longrightarrow \left\{\begin{array}{l}\text{新しい会話状態}\\ \text{応答文}\end{array}\right\}$$

の形の推移規則に従って行う。すなわち，

- 現在の会話状態：現在，７つの状態のどこにいるか，
- 予約内容の把握状態：まだわからない予約項目はどれか，認識結果が一意に決まらないであいまいな予約項目はどれか，全部の予約項目が決まったか，予約項目に矛盾はないか，
- 認識結果：システムからの確認の応答に対して入力の認識結果が「はい」であるか，「いいえ」であるか，予約内容を訂正した入力が認識できたか，リジェクトされたか，

に従って

・新しい会話状態：7つの状態のいずれか，

に推移して

・応答文：新しい会話状態に対応する応答文のいずれかで，予約内容の把握状態によって決まる，

を出力する。

質問回答のやりとりが，できるだけ自然な形で行えるように，会話状態間の推移規則は次の原則に従って作成した。

(1) 予約内容にあいまいな予約項目があるときには，まず，その予約項目の確認を行って，あいまいさをなくすようにする。その際，認識結果の第1位の候補から順に確認し，最後の候補まできたときには，確認しないで，それに決める。なお，認識結果の候補の数には特に制限をもうけていなくて，認識の際に音韻変形規則を適用したことに伴うペナルティの総和が閾値をこえないものすべてを認識結果とし，そのペナルティの総和の少ないものから順に，第1位，第2位，…と順序づける。

(2) 次に，まだわからない予約項目の問合せを行う。

(3) 予約内容に，あいまいな予約項目やまだわからない予約項目が2つ以上ある場合には，日付，発駅，着駅，発時刻，列車名，等，枚数の順に問合せや確認を行う。

(4) 予約項目間の矛盾のチエックは，全部の予約項目が決まってから行い，そこで矛盾が見つかれば訂正を求める。

基本的な原則は以上であるが，質問回答のやりとりができるだけ少ない回数で済むように，次の考慮も払った。

(5) 発時刻と列車名の予約項目にあいまいさがあるときには，すぐには確認を行わないで，次のようにする。発駅が決まると時刻表を用いて発時刻と列車名の候補をかなり強力にしぼることができるので，まず，発駅の予約項目を認識し，それに基づく推論によって，時刻表に合致しない発時刻と列車名を消して，まだあいまいさが残っていれば確認を行う。発時刻と列車名は，他の予約項目に比べて，認識結果が一意に決まらないことが特に多いので，このような推論によって，確認の回数をかなり減らす効果がある。

(6) 通常は，推論を行いながら，予約内容を認識する。しかしながら，認識誤りが生じたときは，それに基づく誤った推論が行われて，認識誤りの訂正ができないことがありうる。このような事態をさけるために，推論によってある予約項目の認識結果が全部消された場合は，誤った推論が行われたかも知れないと判断して，そのあと全予約項目の確認を行うまで，推論の機能を一時止める。

時刻表を用いた推論は，発駅，発時刻，列車名の予約項目の間で行っており，その内容を表8.8に示す。推論が実行されて，予約内容が変ったときには，再び推論ができるかどうか調べて，これを予約内容が変らなくなるまでくり返す。

表8.8 推 論 の 種 類

	推論の型および予約項目の状態	他の予約項目の状態
1	発 駅 (definite) ⇒ 発時刻 (ambiguous)	列車名 (not definite)
2	発 駅 (definite) ⇒ 列車名 (ambiguous)	発時刻 (unknown)
3	発時刻 (definite) ⇒ 発 駅 (ambiguous)	列車名 (unknown)
4	発時刻 (definite) ⇒ 列車名 (ambiguous)	発 駅 (not definite)
5	列車名 (definite) ⇒ 発時刻 (ambiguous)	発 駅 (not definite)
6	発 駅 (definite) 発時刻 (definite) ⇒ 列車名 (not definite)	
7	発 駅 (definite) 列車名 (definite) ⇒ 発時刻 (not definite)	
8	発時刻 (definite) 列車名 (definite) ⇒ 発 駅 (not definite)	

definite ：一意に予約内容が定まっている。
ambiguous ：予約内容があいまいである。
unknown ：予約内容が未定である。
not definite ：予約内容が未定あるいはあいまいである。

次に，上に述べた会話モデルによって質問回答を行った会話例を示す。

（例1）

計算機：（予約開始の案内）

発声者：姫路駅より，小郡駅までの，グリーン券で，5日の，こだま217号を，予約します。

計算機：3日ですか？

発声者：いいえ。

計算機：何枚ですか？

発声者：5枚，です。

計算機：あなたの予約は，5日のこだま217号で，姫路駅8時27分発，小郡駅までのグリーン券を，5枚ですね。

発声者：はい。

計算機：（予約終了の案内）

この例では，日付は，3日（第1位）と5日（第2位）に認識されたが，3日でないことを確認して，5日に決まった。そのあと，まだ入力していない枚数の問合せが行われた。また，発時刻は推論によって決った。

（例2）

計算機：（予約開始の案内）

発声者：新横浜から，浜松への，28日の，こだま215号の，グリーンを，1枚。

計算機：普通席，グリーン席のどちらですか？

発声者：グリーン席。

計算機：御希望の予約は，こだま215号，新横浜駅9時2分発，浜松駅までと了解しましたが該当する列車がございません。当方の聞き違いと思いますので訂正をお願い致します。

発声者：8時34分発。

計算機：あなたの予約は，28日のこだま215号で，新横浜8時34分発，浜松駅までのグリーン券を，1枚ですね？

発声者：そうです。

計算機：（予約終了の案内）

この例では，最初　入力した「グリーンを」を9時2分に誤認識した。そのために，普通席かグリーン席かの問合せが行われ，全部の予約項目が決ったところで，予約項目間の矛盾による訂正が求められた。

8.6　音声応答[94][136]

各会話状態に対応した応答文を，音声合成して出力する。この音声出力は，単語または文節を単位とする録音編集方式で作る。

応答文には，表8.7に示されるように，9種類の固定文と14種類の可変文がある。表8.7で下線を引いたところが可変部分であり，そこに挿入する単語または文節は合計460種類である。そのうちわけは，次の通りである。

日　付（31）：1日～31日
駅　名（28）：東京駅～博多駅
時　間（78）：6時～23時
　　0分発～59分発
列車名（3）：ひかり，こだま
列車番号（294）：1号～195号（ひかり）
　　200～298号（こだま）
券　（2）：普通券，グリーン券
枚　数（9）：1枚～9枚
予約番号（10）：4番，13番，…，811番
その他（6）：何日で，どこから，どこまでで，何時何分発で，どの列車で，
　　普通席グリーン席のどちらで

録音編集のための音声データは，1名の女性が発声した，すべての応答文に必要な文章，文節，単語を，3.2kHzの低域通過のフイルタに通して6.4kHzで標本化し，8ビットに量子化してディスクに蓄えた。ディスクの使用領域は2.3M語（1語16ビット）である。

8.7　認識実験による評価

8.7.1　音響処理[92][133][134]

会話音声認識システムの性能を調べるため認識実験を行った。発声は男性8名（RN, SF, SS, KI, KS, HN, SA, MK）で，入力は文節ごとに区切って発声された会話音声である。各発声者ごとに248文節，計1984文節を資料として用いた。発声リストを付録5に示す。発声者は騒音レベル69 dB(A)の計算機室で行った。用意した標準パターンは5母音と撥音/N/の6種類，VCV音節標準パターンが462種類，文頭の子音を認識するのに必要なCV音節の標準パターン

が76種類である。音韻別のうちわけを表8.9に示す。母音と撥音の標準パターンは各発声者ごとに作成した。VCV音節標準パターン（CV型を含む）は発声者RNのものを全発声者に共通に用いた。

表8.9　標準パターンの種類

母音	VCV 音節					CV 音節				
	*	w	y, d, zy, ky, hy	m, n, b, g, r, z, p, t, k, s, h	計	*	w	y, d, zy, ky, hy	m, n, b, g, r, z, p, t, k, s, h	計
6	36	6	各18	各30	462	5	1	各3	各5	76

音韻ラティス中に正しい音韻が含まれる割合を8名の発声者の平均について求めたものを表8.10に示す。母音は85.9%，子音は71.4%，全体で78.6%である。このうち，母音およびVCV音節標準パターンの学習を行った発声者（RN）の結果は85.1%であり，8名の平均値より6.5%良い値が得られている。

表8.10 音韻ラティス中に正しい音韻が含まれる割合

	正	認識誤り	セグメント化誤り
母音	85.9%	6.4%	7.7%
子音	71.4%	21.2%	7.4%
全体	78.6%	13.8%	7.6%

母音のconfusion matrixを表8.11に示す。/i/，/u/の脱落誤りが多いが，これは無声子音に前後をはさまれた/i/，/u/が無声化することによるものが大部分である。/N/の脱落誤りは，/iN/，/uN/のようにスペクトルの似た母音の後に/N/が続く場合に長い/i/や/u/として認識されることが多いためである。また，/u/が/o/に誤認識される場合の大半は，無声子音の後の/u/が変形しやすく，誤認識されやすいのが原因である。その他の誤りは比較的少ない。

表 8.11　母音の confusion matrix

入力＼結果	a	i	u	e	o	N	脱落
a	1670			8	10		16
i		1152	37	44	1	5	201
u	10	12	596	19	122	3	110
e	11	14	2	509	1		7
o	23	1	9		1152	4	11
N	1	16	5	3	7	293	75
挿入	7	17	5	9	35	3	

次に，子音の音韻別の認識結果を図8.14に示す。有声子音では，半母音 /w/，/y/の認識が悪い。

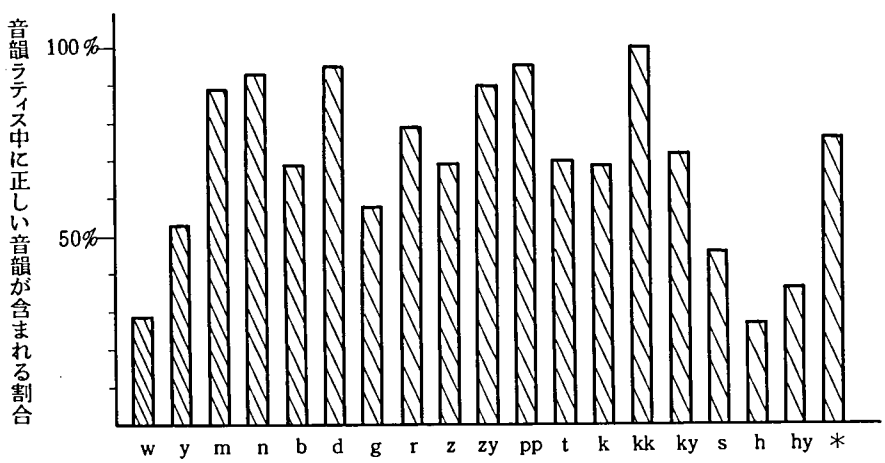


図 8.14　子音別の認識結果

これらの有声子音は母音性の音であり，前後の母音の影響を受けやすいので，連続音声中で存在を検出することが困難なためである。特に，/yoyaku/(予約）のように半母音が連続した場合の検出が困難である。次に結果が悪いのは /g/である。これは，/g/が連続音声中で鼻音化して /m/，/n/に誤認識されやすいのが原因である。無声子音において結果が悪いのは /s/，/h/，/hy/である。これらのうち，/s/，/h/は /sitei/（指定），/hikari/（ひかり）のように

文頭の母音が無声化しやすく，それに伴って脱落することが多いためである。さらに /h/は，前後の母音の影響を受けて母音化しやすいために，脱落誤りが生じやすい。/hy/は /h/と同様母音化しやすく，さらに，/zy/，/ky/等の拗音に比較すると変化が明確でないため，脱落誤りが生じやすい。これらの音韻を除けば安定した結果が得られる。

音響処理の性能をはかる別の基準として，音韻ラティスの持つ情報量から等価的な音韻認識率を推定する方法がある。[148] この方法で求めた音韻認識率を表 8.12 に示す。この場合は，母音が 77.8%，子音が 39.1%，全体では 58.5%の音韻認識率となる。表 8.9 の結果に比較して子音の認識率の低下が大きい。これは，子音の処理が母音に比較して困難なため，子音部分で，音韻の候補が複数個になったり，余分なセグメントの付加が生じやすいのが原因である。

表 8.12　情報量から求めた音韻認識率

	正	認識誤り	セグメント化誤り
母　音	77.8%	2.9%	19.3%
子　音	39.1%	28.1%	32.8%
全　体	58.5%	15.5%	26.0%

8.7.2　言語処理[93][137]

8.7.1 と同じデータを用いて言語処理を行った。発声者ごとの音韻認識率と文節認識率を表 8.13 に示す。文節認識率は，key word の正誤から求めた。平均で 86.0%の文節認識率が得ら

表 8.13　発声者ごとの音韻認識率と文節認識率

(%)

発声者			RN	HN	SF	KI	KS	SS	SA	MK	平均
文節認識率	正 (correct)		95.7	82.2	89.1	90.0	85.2	88.3	74.3	83.0	86.0
	拒絶 (reject)		3.0	13.0	8.7	7.4	11.7	9.6	17.4	10.0	10.1
	誤り (error)		1.3	4.8	2.2	2.6	3.1	2.1	8.3	7.0	3.9
音韻認識率	ラティスに正しい音韻が含まれている割合		85.1	74.5	78.9	78.8	76.4	79.3	77.5	78.8	78.6
	情報量から推定	平均	62.7	55.2	60.6	59.1	58.2	59.3	55.6	57.1	58.5
		母音	80.7	76.0	80.0	78.6	76.4	80.2	74.8	75.9	77.9
		子音	44.7	34.4	41.2	39.6	39.9	38.4	35.0	38.3	39.0

母音の標準パタンは発声者ごとに，子音の標準パタンは発声者RNから作成された。

れた。次に，予約項目ごとの文節認識率を表8.14に示す。構文の複雑な発時刻，列車名の認識率が若干低いことがわかる。誤りのほとんどは，予約項目間の誤り（例：グリーン券を→9時2分を）である。not unique（認識結果が複数あり，その中に正しい結果が含まれている）の場合において，第1位の候補が正しい割合は78.1%であり，システムからの確認の質問文に，発声者が「はい」「いいえ」で答えることによって正しい結果が容易に得られる。

表8.14　予約項目別の認識結果

(%)

予約項目		日付	発駅	着駅	発時刻	列車名	等	枚数	全体
correct	total	87.1	87.5	83.9	80.2	80.3	89.7	90.0	86.0
	unique	37.9	70.0	68.1	56.8	55.7	75.0	63.3	62.5
	not unique (*)	49.2 (73.0)	17.5 (87.5)	15.8 (95.8)	23.4 (46.7)	22.6 (87.2)	14.7 (66.0)	26.7 (89.1)	23.5 (78.1)
reject		5.9	10.9	11.8	13.5	14.9	8.1	7.9	10.1
error		7.0	1.6	4.3	6.3	4.8	2.2	2.1	3.9

(*)はnot uniqueのうちで第1位の候補が正しい割合

次に，単語認識の様子を平均的な認識率を示す発声者KSで調べた。この結果を表8.15に示す。正しく認識された単語の割合は87.2%であり，ほぼ文節認識率と同じ値である。これは，

表8.15 単語の認識結果

文章数	20
文節数	124
単語数	423
正しく認識された単語	369 (87.2%)
誤って認識された単語	767

（発声者　KS）

マッチングがleft－to－rightであることと，意味内容に影響を与えない助詞の誤り(/o/→/no/,

/yuki/—/iki/）があることなどによる。また，正しい単語数の約2.7倍の単語が認識され，そのうち，約1.8倍の単語は，正しくない単語である。この1.8倍の数字は，単語認識での無駄な処理量を表わしており，処理速度をあげるには，認識論理を精密にし，この数字を小さくする必要がある。

8.7.3 認識システムの評価[137]

音韻認識率と文節認識率の関係を調べるために，8名の発声者ごと（1名につき124文節発声したのを2回ないし3回）の音韻認識率と文節認識率をプロットしたのが図8.15である。この図からシステムが良好に動作する目やすである文節認識率95%を達成するには，63%以上の音韻認識率を得る必要があると推定される。

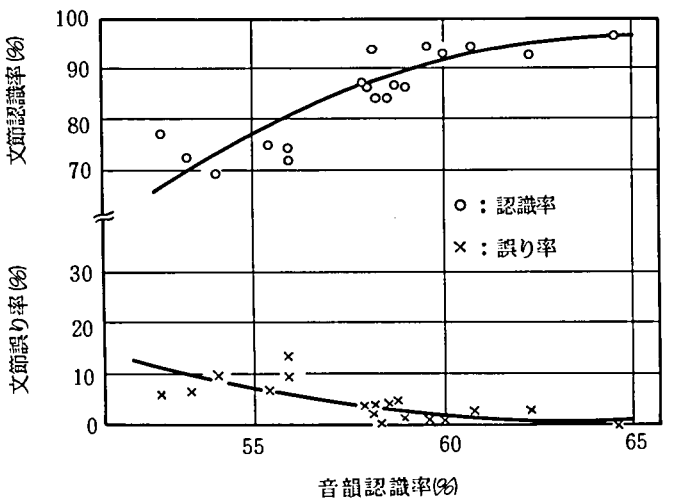
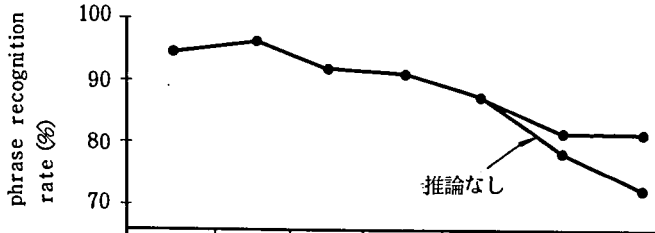


図8.15 音韻認識率と文節認識率の関係

次に，認識対象の複雑さと文節認識率との関係を考察するために，予約項目間にわたる誤りを取り除いたときの文節認識率を図8.16に示す。この図には，構文の複雑さ，意味内容の数，平均の単語数，平均の音韻系列の長さを示した。文節認識率は，構文が複雑である項目ほど，また，音韻系列が長い項目ほど低くなっていることがわかる。



	等	枚数	日付	発駅	着駅	列車名	発時刻
構文の複雑さ＊	81	81	333	1512	2016	3402	50616
意味内容の数	2	9	31	28	28	189	1020
単語数の平均	2.8	2.1	2.9	2.4	2.9	6.1	6.7
音韻数の平均	10.8	8.1	9.6	12.1	13.2	18.8	19.5

＊構文表現が表わす異なる単語列の数

図 8.16 予約項目の複雑さと文節認識率

音韻処理部での音韻認識の誤りと，言語処理部で訂正できなかった誤りとを対比して表8.16に示す。ただし，この表では，余分なセグメントの付加も誤りとしてかぞえた。音響処理部で多

表 8.16 音響処理での音韻認識の誤りと言語処理で訂正できなかった誤り

音響処理部での誤り			言語処理部での誤り		誤り訂正率 $\frac{(A-B)}{A}\times 100(\%)$
割合(%)	個数 A	誤りの原因	個数 B	割合(%)	
2.3	104	文頭の母音の誤り	32	21.4	69.2
2.5	117	拗音のセグメント化誤り	29	19.5	75.2
1.6	71	子音過渡区間→母音	16	10.7	77.5
2.5	116	撥音 /N/の認識誤り	16	10.7	86.2
8.8	406	無声子母の後の母音	8	5.4	98.0
1.8	84	連続母音	6	4.0	92.9
7.2	334	その他の母音の調音結合	12	8.1	96.4
0.9	41	子音の挿入	6	4.0	85.4
22.4	1038	子音間の認識誤り	5	3.4	99.5
0.4	20	母音→子音	3	2.0	85.0
0.2	10	ポーズ→無声子音	1	0.7	90.0
0.0	1	2重の無声化	1	0.7	0.0
43.1	1996	余分なセグメントの付加	9	6.0	99.5
6.3	290	子音の脱落	0	0.0	100.0
—	—	プラグマティクス	5	3.4	
	4628	合計	149		96.8

(発声者8人 各自124文節)

い誤りは，余分なセグメントの付加と子音認識の誤りであるが，これらは言語処理部で大部分訂正できる。したがって，会話音声認識システム全体での性能向上という観点からは，言語処理での誤りの原因となっている音響処理の誤りに重点をおいて検討を行う必要がある。それらは次の4項目である。

(1) 文頭の母音の誤り

(2) 拗音のセグメント化の誤り

(3) 有声子音や過渡的な音を母音に間違える誤り

(4) 撥音 /N/のセグメント化誤りと認識誤り

今後は，これらの誤りを少なくすることに重点をおいて音響処理の改善をはかる必要がある。

第7章で述べた第1次システムと，本章で述べた第2次システムの比較を表8.17に示す。

表 8.17 第1次システムと第2次システムの比較

			第1次システム	第2次システム
対象	タスク		新幹線の座席予約サービス	
	予約項目		6項目	7項目（日付を追加）
	駅名		13駅（東京～新大阪）	28駅（東京～博多）
	列車本数		56本	181本（通常の時刻表）
	発声の制限		文節ごとに0.5秒以上のポーズ	
音響処理部	母音の標準パターン		6種類（/a/, /i/, /u/, /e/, /o/, /N/），最尤パラメータ	
	子母の標準パターン		468種類（VCV型）	539種類（VCV型＋CV型）
	音韻規則		なし	9型，48種，top－down的利用
	相関分析		CPU	実時間相関器（15 msecごと）
	母音認識		CPU，最尤法，bottom－up	HSSP，最尤法，bottom－up
	VCV音節認識		CPU 端点固定PP bottom－up的に利用	HSSP，端点フリーDP， top－down的に利用
	処理結果		音韻ラティス	音韻ラティス＋継続時間
言語処理部	音韻変形規則		8型，165種 top－down的利用	14型，477種，top－down的利用 連想的な検索
	単語辞書		72単語，音素表記	112単語，音素表記
	構文表現		文節ごとに手続きとしてリスト表現 left－to－right	
	プラグマティクス		プログラムで記述	知識としてリスト表現で記述
	単語認識		depth－first，top－down， left－to－right，ペナルティの制限，処理済みの状況の利用	depth－first，top－down， left－to－right，ペナルティの制限 fall－back system，処理済みの状況を連想的に利用
	構文解析		top－down，depth－first， left－to－right	top－down，depth－first，left－to－right マッチングの度合の利用
計算機	音響処理	CPU	FACOM 270/20 コア 32kw，サイクルタイム2.4 μs	NEAC 3200/70 コア 32kw，サイクルタイム800ns
		主な使用言語	フォートラン	フォートラン
	言語処理	CPU	FACOM 270/20 コア 32kw，サイクルタイム2.4 μs	パナファコム U－400 コア 32kw，サイクルタイム800ns
		主な使用言語	DLOP（フォートラン）	DLOP（フォートラン＋アセンブラ）
	音響・言語の結合		オフライン	オンライン，並列処理
処理時間	音響処理		実時間の80倍	実時間の3.4倍
	言語処理		〃 15倍	〃 2.0倍
	全体		〃 95倍	〃 5.0倍
性能	発声者		男性1名	男性8名（母音学習，VCV共通）
	発声環境		防音室	計算機室（69dB(A)）
	音韻認識率		62.1%	58.5%
	文節認識率		87%	86%

8.7.4 標準パターンの検討[138]

以上述べた認識実験では，標準パターンとして，母音は個人別のものを，VCV音節は共通のものを用いた。ここでは，さらに，すべての標準パターンを個人別に作った場合と共通した場合の認識実験を行い，結果を比較検討することによって，システムの性能と標準パターンの関係を調べる。

8名の男性発声者（RN, SS, MK, SF, KI, KS, SA, HN）が発声した各自124文節，計992文節を認識実験に用いた。実験条件としては，言語処理部は固定しておき，音響処理部における母音標準パターン（撥音を含む），VCV音節標準パターンを表8.18に示す3通りの条件にして認識実験を行った。

表8.18 認識実験の条件

	母音標準パターン	VCV音節 標準パターン
条件 I	共　　通	共　　通
条件 II	個　人　別	共　　通
条件 III	個　人　別	個　人　別

ただし，共通の標準パターンとしては，次に示すものを用いた。

母音標準パターン……本人を除いた7名の標準パターンの平均

VCV音節標準パターン……発声者RNの標準パターン

音響処理，言語処理を行って得られた文節認識結果を表8.19に示す。また，図8.17には個人別の文節認識率を，図8.18には予約項目別の文節認識率を示す。これらから次のことがわかる。

表8.19 文節認識結果（かっこの中は正しい結果が1位になるものの割合）

	correct			reject	error
	total	unique	not unique		
条 件 I	80.0%	56.5%	23.5% (75.9%)	15.0%	5.0%
条 件 II	86.7%	62.0%	24.7% (78.0%)	10.6%	2.7%
条 件 III	86.2%	63.0%	23.2% (86.9%)	11.3%	2.5%

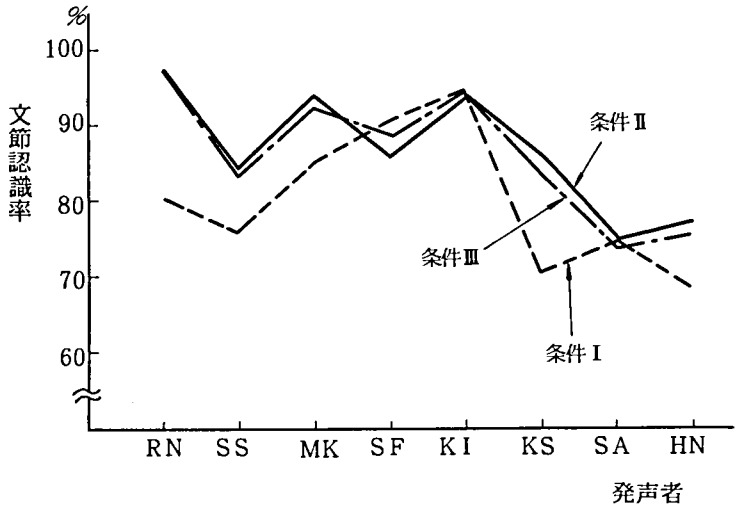


図 8.17　個人別の文節認識率

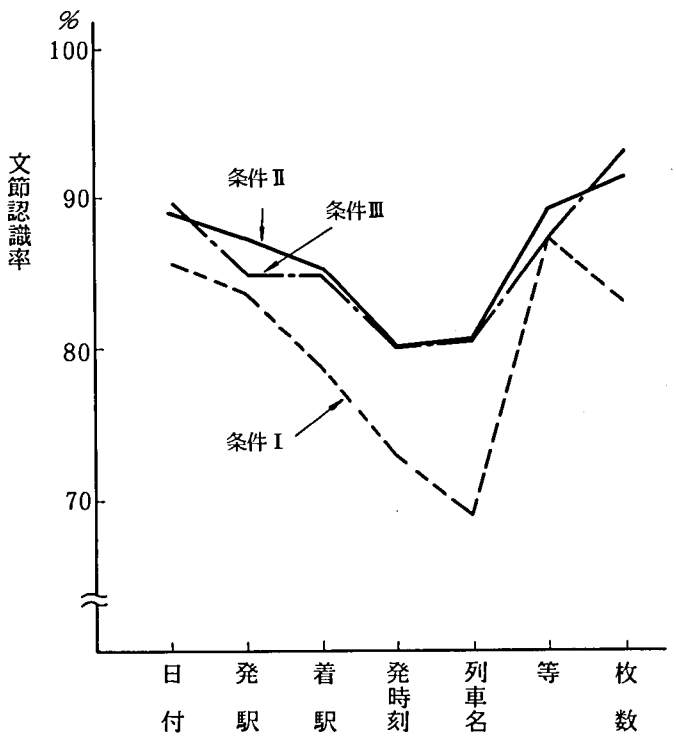


図 8.18　予約項目別の文節認識率

(1) 条件Ⅰ，Ⅱの結果を比較すると，平均の母音標準パターンを用いると，文節認識率が6.7％低下することがわかる。また，認識誤りもほぼ2倍に増加している。このことは，母音標準パターンを発声者ごとに学習する方式が有効であることを示している。項目別に見ると，平均標準パターンを用いる方法では，発時刻，列車名など，長い文節やセグメント誤りの生じやすい文節で認識率の低下が大きい。

(2) 条件Ⅱ，Ⅲの結果を比較すると，表8.17，図8.17，図8.18のいずれにおいてもほとんど同じ結果が得られている。これは，システム全体の性能への子音認識結果の影響が母音に比較して小さいことを示している。このことから，VCV音節標準パターンは特定の発声者のものを共通に用いても良いといえる。

8.8 質問回答実験による性能評価[94][136]

8名の男性発声者を用いて，7項目全部の予約が完了するまで質問回答をくり返す実験をそれぞれ40種類の予約について行った。ただし，同じ内容の質問回答が5回続く場合は打ち切ることとした。また，各発声者の結果を同じ条件で比較するため，各予約における最初の発声については，発声内容をあらかじめ定めておいた。発声は騒音レベル69 dB(A)の計算機室で行った。次に，いくつかの面からシステムの評価を行う。

8.8.1 予約完了率

最後まで質問回答が行われ，予約が完了した割合を表8.20に示す。全体での完了率は99.1％である。質問回答形式をとることにより，きわめて高い完了率が得られることがわかる。打ち切りの原因を調べると，

(1) 発声者SFの場合は，「浜松から」において鼻音の前後の母音/a/が鼻音化して/o/になったため言語処理レベルで誤りの訂正ができなかったもの，

表8.20 予 約 完 了 率

発 声 者	RN	HN	SF	KI	KS	SS	SA	MK	TOTAL
打ち切りの数	0	0	1	0	0	2	0	0	3
予約完了率	100％	100％	97.5％	100％	100％	95.0％	100％	100％	99.1％

(2) 発声者SSの場合は，「6枚」において，先頭の子音/r/が強く発声されるため，母音/u/に誤り，その訂正ができなかったもの，

である。これらの原因は，音響処理，言語処理レベルの改善によって比較的容易に除去できる。

8.8.2 質問回答の回数

途中で質問回答を打ち切った3個を除いて，予約完了までの質問回答の回数を個人別の表にして表8.21に示す。表8.21には最初の発声における文節認識率もあわせて示した。質問回答

表8.21 質問回答の回数

発声者	内容に関する質問回答			確認に関する質問回答		計	文節認識率
	必要な最小の回数*	いいなおし	誤り訂正	「はい」	「いいえ」		
R N		0.18	0.00	1.78	0.26	3.70	95.2%
H N		0.34	0.20	1.35	0.48	3.85	87.7%
S F		0.52	0.49	1.51	0.18	4.18	87.7%
K I	1.48	0.37	0.43	1.55	0.50	4.33	88.5%
K S		0.57	0.43	1.75	0.50	4.73	85.0%
S S		0.63	0.47	1.74	0.63	4.95	78.4%
S A		1.46	0.33	1.80	0.53	5.60	78.0%
M K		0.79	0.83	2.23	0.90	6.23	76.7%
	1.48	0.61	0.39	1.71	0.50	4.69	84.6%

*完全に確認できた場合は，1.48回の質問回答で予約が完了する。

の内容は，(1)予約内容に関する発声，(2)確認を求める応答文（「…ですか？」）に対する「はい」「いいえ」の発声に分けられる。(1)はさらに，各予約項目に関する最初の発声，発声した内容がリジェクトされたため発声しなおしたもの，発声した内容が誤って認識されたため，それを訂正するための発声，に分けられる。全体での平均の質問回答回数は4.69回である。このうち，1.48回は予約が完了するために必要な最小の質問回答回数である。これが1より大きいのは，最初に必ずしも全項目を発声しないためである。残りの内わけは，いいなおしが0.61回，誤り訂正が0.39回，「はい」が1.71回，「いいえ」が0.50回である。「はい」の回数が多いが，こ

れは発声者にあまり負担はかけない。これに対し，発声者に対する負担が大きいと思われる誤り訂正の回数は少ない。したがって，本システムは発声者にとって使いやすいシステムであるといえる。

次に，発声者別に見ると，質問回答の回数は3.70回から6.23回まで分布している。また，文節認識率は95.2％から76.7％まで分布している。質問回答の回数と文節認識率の間には強い相関があり，文節認識率が悪くなるに従って，質問回答の回数が増加する。しかしながら，発声者MKは発声者RNに比較して，文節認識率が20％近く低下しているにもかかわらず，必要な質問回答の回数は約2.5回増加しているにすぎない。この程度の質問回答数の増加は，発声者にとって大きな負担になるとは考えられない。したがって，質問回答形式で認識を行うことの有効性がわかる。

8.8.3 項目別の発声回数

表8.22に各項目別の発声回数を示す。表において，値が1より大きいのは，リジェクトによるいいなおし，誤り訂正があることを示している。発時刻，列車名は，数字が多く含まれ，しかも連続した単語数が多いので，他の項目に比べて認識が困難である。それにもかかわらず，発声回数が他の項目に比較して少ないのは，時刻表を用いた推論に負うところが多い。特に，列車名の発声回数は平均0.99回であって，推論によって，発声しなくても良い場合があること

表8.22 項目別の発声回数

項目 / 発声者	日付	発駅	着駅	発時刻	列車名	等	枚数	平均
RN	1.05	1.08	1.02	0.73	0.88	1.00	1.03	0.97
HN	1.10	1.10	1.13	0.83	1.00	1.08	1.08	1.04
SF	1.03	1.26	1.23	1.28	1.03	1.13	1.08	1.15
KI	1.08	1.08	1.13	0.93	0.93	1.18	1.23	1.08
KS	1.13	1.13	1.25	1.03	1.05	1.18	1.15	1.12
SS	1.16	1.18	1.32	1.13	1.00	1.18	1.13	1.16
SA	1.10	1.53	1.78	1.43	1.00	1.23	1.13	1.31
MK	1.30	1.20	1.23	1.33	1.03	1.30	1.45	1.26
平均	1.12	1.19	1.26	1.08	0.99	1.16	1.16	1.14

を示している。質問回答の回数と，項目別の発声回数の発声者間の順位は必ずしも一致しない。これは，発声者によって認識のされかたに差があること，すなわち，あいまいに認識されやすい発声者や，あいまいさは少ないがジェクトされやすい発声者が存在することを示している。

8.9 あとがき

列車の座席予約を対象とした会話音声認識システムの第2次システムについて，音響処理部の構成および性能を中心に述べた。本音響処理の特徴は次の3点である。

(1) 音声の分析を最尤スペクトル分析で行うことにより，精密なスペクトル情報の抽出をはかった。

(2) 従来の音響処理がbottom－up的に行われていたのに対し，音韻認識をtop－down的に行う方法をとり入れた。

(3) 必要な標準パターンのうち，比較的学習の容易な母音標準パターンのみを発声者ごとに学習する方式をとり，発声者に対する適応性を高めた。

男性8名を用いた認識実験の結果，音韻ラティス中に正しい音韻が含まれる割合は78.6％であった。言語処理をあわせた会話音声認識システム全体では，8名の発声者の平均で86.0％という文節認識率が得られた。また，VCV音節標準パターンを個人別に用意しても文節認識率はさほど変化しないことを示し，母音標準パターンのみを発声者ごとに学習せる方式が有効であることを示した。

同様に，男性8名によって質問回答形式で予約を行う実験を行った。その結果，平均4.69回の質問回答で99.1％という高い予約完了率が達成できることを示した。

第9章　音韻を単位とした単語音声の認識

9.1　はしがき

前章まではVCV音節を単位とした音声認識法について述べた。本章および次章では音韻を単位とした音声認識法について述べる。

VCV音節を単位とした音声認識法は一般的な連続音声の音響処理を目標としたものである。これに対し，数十～数百の語彙を対象とした単語音声，連続単語音声といった認識対象に対して実用に供しうるような性能を得るには，認識対象に合致した方法をとる必要がある。音韻を単位とした認識方法はそのような立場から開発した手法である。

従来，単語音声認識を行う場合には，単語単位の標準パターンを用いる方法が多く用いられていた。この方法は，発声された単語そのものを標準パターンとするために調音結合の問題を回避できるので，比較的容易に高い認識率を得ることができる。しかしながら，標準パターンを登録する時に認識対象のすべての単語を発声しなければならず，発声者に大きな負担がかかる。また，標準パターンを蓄えておく記憶容量が膨大になるなどの欠点がある。したがって，語彙数を増大させようとする場合に問題が多い。これに対し，本章で述べる認識方法は，標準パターンを音韻単位で蓄えておき，一方，認識対象の単語は音韻系列の形で蓄えておき，これらを併用することにより単語音声を認識する方法であり，次のような特徴を有している。

(1)　標準パターンと辞書とを独立して有しているため，認識対象の単語を変更する場合には単語辞書の内容を変更するだけでよい。

(2)　音韻標準パターンを登録する場合には，必ずしも全部の単語を発声する必要はない。一部の単語を発声するだけで標準パターンの登録を行うことができる。[52][53]

(3)　単語を標準パターンとして用いる方式に比較して，標準パターン，単語辞書を蓄えておくための記憶容量が少なくてすむ。

入力音声と辞書の音韻系列の照合を行う際に，発声に伴って生じる時間軸の非線形な伸縮を補正してマッチングを行うため，DPマッチングの手法を用いる。その際，2種類のDPマッチング方法を提案し，認識実験により方法の評価を行う。

9.2 単語音声認識系の構成[139]

認識系は図 9.1 に示すように，前処理部，類似度計算部，DPマッチング部から構成されている。

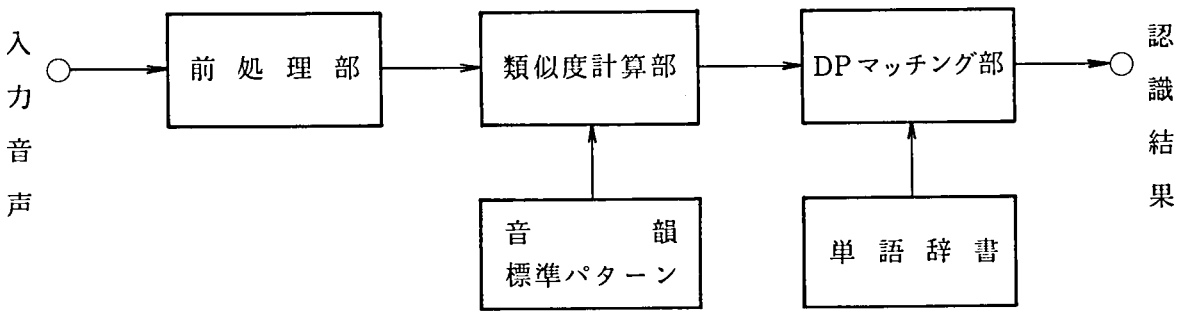


図 9.1 単語音声認識系の構成

9.2.1 前処理部

前処理部の構成を図 9.2 に示す。入力音声は，まず，6dB/oct の高域強調をかけ，3.2kHz の低域通過フィルタを通した後，サンプリング周波数 8 kHz で 11 ビットに量子化される。ディ

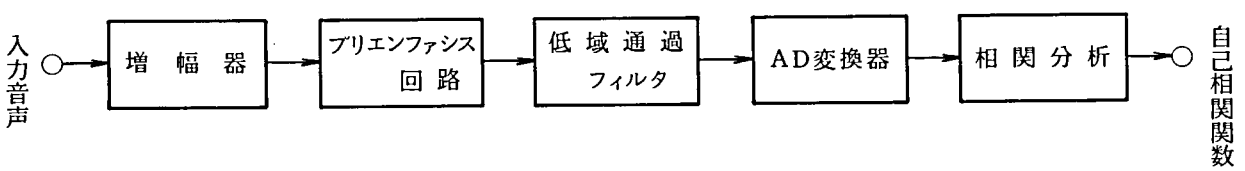


図 9.2 前処理部の構成

ジタル化された音声は 15 msec のフレームごとにパワーが求められる。あらかじめ定められた閾値との比較により音声区間を抽出する。パワーが閾値より大きくなったフレームを音声区間の始端とする。閾値以下のパワーが 20 フレーム以上続いた場合は，最初に閾値以下になったフレームを音声区間の終端とする。音声中に閾値以下のパワーが 20 フレーム以下続く場合は，無声子音等の前で生じる無音区間と見なす。抽出された音声区間では，15 msec のフレームごとに音声の特徴パラメータとして，音声波形の自己相関関数

$$\boldsymbol{v} = (v_0, v_1, \cdots\cdots, v_p) \qquad (9.1)$$

が求められる。したがって，入力音声は自己相関関数の時系列

$$V = (v_1, v_2, \cdots\cdots, v_N) \qquad (9.2)$$

として求められる。ただし，

$$v_i = (v_{i0}, v_{i1}, \cdots\cdots, v_{ip}) \qquad (9.3)$$

は第 i フレームの自己相関関数である。

9.2.2 類似度計算部

類似度計算部では，入力音声の各フレームの特徴パラメータと，音韻標準パラメータとの類似度を求め類似度マトリクスを作成する。音韻標準パターンは最尤スペクトルパラメータの形で蓄えられている。音韻 /x/の標準パターンの最尤スペクトルパラメータを

$$A_x = (A_{x0}, A_{x1}, \cdots\cdots, A_{xp}) \qquad (9.4)$$

とする。第 i フレームの自己相関関数，式（9.2）と音韻 /x/の標準パターン，式（9.4）との類似度として

$$l(i, x) = -\log\left\{\sum_{\tau=0}^{p} A_{x\tau}\, v_{i\tau}\right\} \qquad (9.5)$$

を用いる。これは第2章で定義したスペクトル間の類似度の式である。

この結果，入力音声は列がフレーム数，行が音韻の種類で，要素を類似度とする行列で表わされることになる。これを類似度マトリクスと呼ぶ。

9.2.3 DPマッチング部

DPマッチング部では，類似度マトリクスと辞書の音韻系列との照合をDPマッチングを用いて行う。DPマッチング法として2種類の方法を提案する。方法Ⅰは，辞書の音韻系列の表現法として，各音韻に平均的な継続時間長（フレーム数）の制限をつけて表現する方式である。これに対して方法Ⅱは，音韻系列の各音韻に最小継続時間と最大継続時間の制限をつける方法である。それぞれの方法について説明する。

9.2.3.1 方法Ⅰ

認識対象の単語は単語辞書中で音韻系列の形で蓄えられている。単語 w^r の音韻系列表示を

$$w^r = x_1\ x_2 \cdots\cdots x_j \cdots\cdots x_J \qquad (9.6)$$

$$(r=1,\ 2,\ \cdots,\ R)$$

とする。式（9.6）は，単語w^rがJ種類の音韻の系列として表現されていることを示している。Rは単語辞書中の音韻系列の数である。各音韻には継続時間情報が付加されている。x_jの平均的な継続時間（フレーム数）をD_jとするとw^rの各音韻をその平均継続時間だけ並べた長さ$\sum_{j=1}^{J} D_j$の系列を用意する。w^rを展開してできるこの音韻系列を，改めて

$$wa^r = y_1,\ y_2 \cdots\cdots y_k \cdots\cdots y_K \qquad (9.7)$$

$$(K = \sum_{j=1}^{J} D_j)$$

とする。w^rとこれを展開してできるwa^rの例を図9.3に示す。

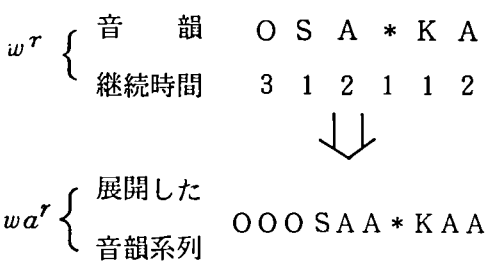


図9.3　音韻系列とその展開形（方法Ⅰ）

入力音声，式（9.1）と音韻系列，式（9.7）の類似度は次式によって与えられるものとする。

$$L(r) = \max_{f} \left\{ \sum_{i=1}^{N} l(i,\ y_{f(i)}) \right\} \qquad (9.8)$$

ただし，$l=(i,\ y_j)$は第iフレームの入力音声と音韻y_jの類似度である。また，fは式(9.1)と式（9.7）の対応づけを行う際の制限を定める関数で，次の関係がある。

$$f(1) = 1, \qquad f(N) = K \qquad (9.9a)$$

$$f(i) = \begin{cases} f(i-1) \\ f(i-1)+1 \\ f(i-1)+2 \end{cases} \qquad (9.9b)$$

$$\frac{K}{N} - Dw \leqq f(i) \leqq \frac{K}{N} + Dw \qquad (9.9c)$$

式（9.9a）はマッチングの始端と終端の条件を示しており，両パターンの始端と終端を一致させてマッチングすることに相当する。式（9.9b）は関数 f の局所的な制限を示している。また，（9.9c）は関数 f の値の範囲に関する制限を示している。この様子を図9.4に示す。すなわち，

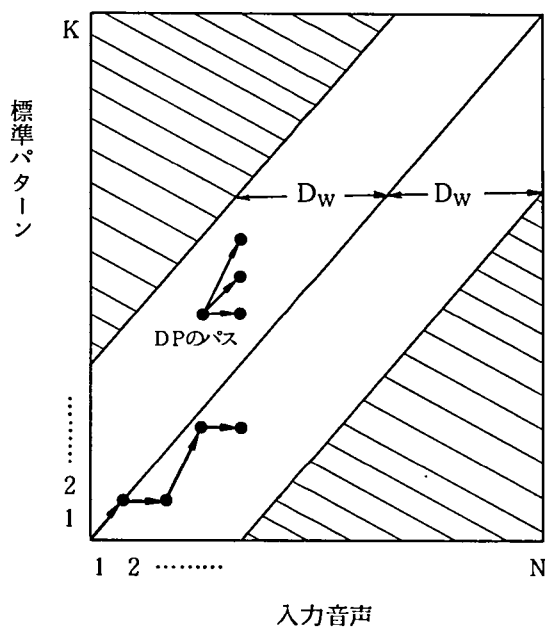


図9.4 方法IによるDPマッチング

入力音声を横軸，音韻系列を縦軸にとったマトリクス上で幅2Dw+1の窓の範囲内で点(1,1)から点（N，K）へ至る曲線を式（9.8）を満足するように求めることに相当する。式（9.8）を求めるのは次のようにDPを用いて効率良く行うことができる。

$$L(i,\ k|r) = \max \begin{cases} L(i-1,\ k|r) + l(i,k) \\ L(i-1,\ k-1|r) + l(i,k) \\ L(i-1,\ k-2|r) + l(i,k) \end{cases} \qquad (9.9)$$

式（9.9）の漸化式を

$$\left.\begin{aligned} &L(i,-1|r) = L(i,-2|r) = L(-1,\ k|r) = -\infty \\ &L(1,\ 1|r) = l(1,\ 1) \end{aligned}\right\}$$

なる初期条件のもとで求めていくと

$$L(r) = L(N, K \mid r) \tag{9.11}$$

となって式（9.8）の$L(r)$が求められる。

$$\max_{r} L(r) \tag{9.12}$$

を満足する単語w^{r*}が認識結果である。

9.2.3.2 方法II

方法Iと同様に単語w^rの音韻系列表示を

$$w^r = x_1\ x_2\ \cdots\cdots\ x_j \cdots\cdots\ x_J \quad (r = 1, 2, \cdots, R) \tag{9.13}$$

とする。各音韻には継続時間情報が付加されている。ただしこの場合，方法Iと異なってx_jには最小継続時間$D_j{}^m$と最大継続時間$D_j{}^M$の制限が付けられている。このとき，入力音声，式（9.1）とw^rのマッチングは

$$\left.\begin{aligned} &1 \leqq n_1 < n_2 \cdots\cdots < n_{J-1} < N \\ &D_j{}^m \leqq n_j - n_{j-1} \leqq D_j{}^M \end{aligned}\right\} \tag{9.14}$$

なる制限のもとで次式を求めることによって得られる。

$$L(r) = \max_{\{nj\}} \left\{ \sum_{n=1}^{n_1} l(n, x_1) + \sum_{n=n_1+1}^{n_2} l(n, x_2) + \cdots + \sum_{n=n_{J-1}+1}^{N} l(n, x_J) \right\} \tag{9.15}$$

ここで$\{nj\}$は音韻の変化時点を示している。式（9.15）の右辺は単語w^rの各音韻の最小，最大の継続時間の制限のもとで，入力音声とw^rの最大類似度を求めることを意味している。この結果得られた$L(r)$が入力音声とw^rの類似度であるとする。

式（9.15）の計算をDPを用いて効率良く行うには次のようにする。まず，w^rの各音韻をその最大継続時間だけ並べた長さ$\sum_{j=1}^{J} D_j{}^M$の系列を用意する。次に，各音韻記号にフラグを設け，x_jの最初の$D_j{}^m$個はフラグを1にし，残りは0にする。この，w^rを展開してできる音韻系列

を改めて

$$wb^{\tau} = y_1\ y_2 \cdots\cdots\ y_k \cdots\cdots\ y_K \qquad (9.16)$$

とする。w^{τ}とwb^{τ}の例を図9.5に示す。このとき入力音声の部分系列$v_1, v_2 \cdots v_i$と音韻系列の

図9.5 音韻系列とその展開形（方法Ⅱ）

部分系列$y_1\ y_2 \cdots\ y_k$の類似度を$L(i, k \mid r)$で表すと，式（9.15）は次の漸化式表現に書き直すことができる。

(1) y_kのフラグが1なら

$$L(i,\ k \mid r) = L(i-1,\ k-1 \mid r) + l(i,\ y_k) \qquad (9.17a)$$

(2) y_kのフラグが0なら

$$L(i,\ k \mid r) = \max \begin{cases} L(i-1,\ k-1 \mid r) + l(i,\ y_k) \\ L(i,\ k-1 \mid r) \end{cases} \qquad (9.17b)$$

ただし，$L(i,\ k \mid r)$の初期値は次のようにおく。

$$\left.\begin{aligned} L(0,\ 0 \mid r) &= 0 \\ L(i,\ 0 \mid r) = L(0,\ k \mid r) &= -\infty\ (i, k \neq 0) \end{aligned}\right\} \qquad (9.18)$$

$L(i,\ 0 \mid r)$を順次計算すると

$$L(r) = L(N,\ K \mid r) \qquad (9.19)$$

となって，式（9.15）の$L(r)$が求められる。図9.6に入力音声と辞書の音韻がマッチングされ

る例を示す。両者の対応が可能なパスを実線で示してある。式（9.17a），（9.17b）の計算はこれらのパスの中で最適のパスを求めることに相当する。

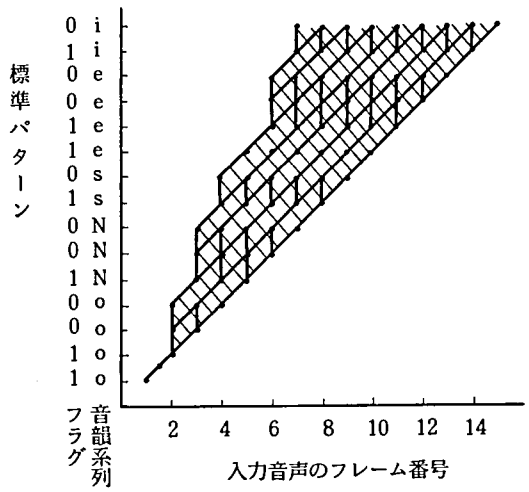


図 9.6　方法ⅡによるDPマッチング

方法Ⅰではマッチング計算を行う範囲を限定するために，窓をかけるという操作を行っており，そのため入力音声終了後でないとDP計算が開始できない欠点があった。それに対し，方法Ⅱでは，DPパスそのものがマッチングの範囲を制限しており，必ずしも窓をかける必要はない。したがって入力音声に同期したDPマッチング計算が可能であり，実時間処理に向いた方法である。

9.3　認識実験[140]

9.3.1　認識対象

この認識系では，数十個の音韻の標準パターンを登録するだけで，100単語程度の認識が可能なので，種々の用途が考えられるが，発声者が認識対象以外の単語を発声した場合とか雑音による認識率の低下等を考慮すると，認識結果がすぐに確認でき，訂正も容易な対話形式での利用が望ましい。ここでは，そのような対象の一例として単語音声による航空機の座席予約サービスを想定し，それに必要な単語を認識対象とした。予約項目は，航空会社名，発空港，着空港，日付，発時刻，枚数，便番号の7項目である。表9.1に項目別の認識対象単語を示す。なお，実際に用いた単語辞書の例を方法Ⅱの場合について付録6に示す。

表9.1 認 識 対 象 の 単 語

	単語数	内　　容
1. 月	12	1月～12月
2. 日	31	1日～31日
3. 時	16	7時～21時
4. 分	12	0分～55分（5分おき）
5. 枚　数	10	1枚～10枚
6. 航空会社	10	日本航空，日航，全日空，東亜国内航空，東亜国内，東亜，日本近距離航空，日本近距離，近距離航空，南西航空
7. 便番号	11	2～3桁の数字と便を区切って発声する。
8. 空港名	67	札幌，旭川，女満別，稚内，釧路，帯広，函館，秋田，青森，八戸，花巻，東京，大阪，南紀白浜，新潟，岡山，隠岐，米子，出雲，徳島，高松，高知，広島，宇部，松山，大分，福岡，宮崎，鹿児島，種子島，屋久島，喜界島，奄美大島，徳島，沖永良部，山形，仙台，八丈島，大島，三宅島，富山，金沢，福井，鳥取，名古屋，北九州，長崎，熊本，福江，佐渡，壱岐，紋別，中標津，沖縄，久米島，南大東，宮古，多良間，石垣，与那国，千歳，小松，大村，那覇，三沢，利尻，奥尻，

9.3.2 音韻標準パターン

日本語の音韻の種類は，音韻論として古くから研究されているが，実際に発声された音声は，調音上の制約や個人差，方言などによる変形が激しいため，いろいろな解釈があり，確実的なものはない。日本語の子音の分類表の例を表9.2に示す。本実験では，表9.2および認識対象

表9.2 日 本 語 の 子 音

調音状態＼調音位置		両唇音	歯音	歯茎音	硬口蓋音	軟口蓋音	声門音
破裂音	無声	p		t		k	
	有声	b		d		g	
通鼻音	無声						
	有声	m		n, η			
摩擦音	無声	F	S		δ		h
	有声	w	z		j		
弾音	無声						
	有声			r			
破擦音	無声		ts				
	有声		dz				

を考慮して次の2種類の音韻標準パターンを用意することとした。

9.3.2.1 音韻標準パターンⅠ

母音は/a/, /i/, /u/, /e/, /o/の5つ，子音は表9.2に示すものと撥音/N/を考え，この中から必要なものだけを使うことにした。認識に用いる音韻は次の23種である。

母　音：/a/, /i/, /u/, /e/, /o/

子　音：/b/, /d/, /g/, /r/, /z/, /w/, /j/,
/m/, /n/, /p/, /t/, /k/, /ts/, /tʃ/, /s/, /h/, /N/

休止記号：＊（語中の無音区間を表わす）

9.3.2.2 音韻標準パターンⅡ

音韻標準パターンⅠに次のものを追加する。

(1) 母音/i/と母音/a/, /u/, /o/の間のわたりを示す標準パターン

(2) 語尾の無声子音の後，および無声子音にはさまれて無声化した/i/, /u/を表わす標準パターン。（記号/i'/, /u'/で示す。）

(3) 無声子音と鼻音の間の/u/を表わす標準パターン。（$/\hat{u}/$）

(4) 音節/ni/において，鼻音/n/のために鼻音化した/i/を表わす標準パターン。（$/\tilde{i}/$）

(5) 音節/ni/において，母音/i/によって口蓋化(注)した/n/の標準パターン。（$/n_i/$）

(6) 子音/k/は，後続の母音によって影響を受けやすいので，それぞれ別の標準パターンとする。（$/k_i/$, $/k_a/$, $/k_o/$, $/k_u/$）

(7) 語中の音節/zi/, /ri/における/z/, /r/は別の標準パターンとする。（$/\overline{z}/$, $/\overline{r}/$）

音韻標準パターンは次の37種である。

母　音：/a/, /i/, /u/, /e/, /o/, $/\tilde{i}/$, /i'/, $/\hat{u}/$, /u'/

母音のわたり：/io/, /iu/, /ai/, /oi/, /ui/

子　音：/b/, /d/, /g/, /r/, $/\overline{r}/$, /z/, $/\overline{z}/$, /w/,
/j/, /m/, $/\overline{n}/$, $/\overline{n_i}/$, /p/, /t/, /k/, $/k_i/$, $/k_a/$
$/k_o/$, $/k_u/$, /ts/, /tʃ/, /s/, /h/, /f/, /N/

注）/i/の音を発声した場合のように，前舌面が硬口蓋に向ってもち上ること。

休止記号：＊

9.3.2.3 音韻標準パターンの作成

音韻標準パターンは，各音韻ごとに1つずつ用意する。標準パターンは学習[19]により，自動的に行われる。ただし，学習の際初期値となる標準パターンが必要である。この初期値となる標準パターンは，あらかじめいくつかの学習サンプルから視察によって各音韻に相当する区間を切り出して作っておく。

9.3.3 実験結果

DPマッチングにおける方法Ⅰ，Ⅱ，および音韻標準パターンⅠ，Ⅱを組み合わせて表9.3に示す3種類の実験を行った。各実験に用いた資料を下に示す。

表9.3 実 験 の 分 類

DPマッチング / 標準パターン	方 法 Ⅰ	方 法 Ⅱ
音韻標準パターンⅠ	実 験 Ⅰ	
音韻標準パターンⅡ	実 験 Ⅱ	実 験 Ⅲ

(1) 実験Ⅰ，Ⅱに用いた資料

- 男性10名が空港名67単語を4回ずつ発声した2680個のサンプル
- 上の男性のうち5名が空港名以外の単語グループを4回ずつ発声した2040個のサンプル。

(2) 実験Ⅲに用いた資料

- 男性12名が便番号を除いた7つの単語グループについて各3名ずつ発声した音声サンプル。

まず，実験Ⅰについて空港名の認識結果を表9.4に，空港名以外の単語の認識結果を表9.5に示す。またこれらを図にして図9.7に示す。さらに，各単語グループごとのconfusion matrixを表9.6～表9.10に示す。

表9.4　認識結果（実験Ⅰ）

対象		空港名
資料の数		670
発声者別の誤りの数	SK	3
	NR	0
	HK	1
	NH	0
	SS	1
	KM	3
	NS	3
	TY	1
	SH	0
	IK	0
TOTAL		12
認識率（%）		98.2

表9.5　認識結果　（実験Ⅰ）

対象		月	日	時	分	枚数	航空会社	便番号	計
資料の数		240	620	320	240	200	200	220	2040
発声者別の誤りの数	HK	1	10	8	3	1	0	0	23
	NH	7	11	9	0	1	0	0	28
	NR	6	13	20	0	4	0	1	44
	SK	4	9	10	1	0	0	4	28
	SS	4	11	11	3	3	0	0	32
	計	22	54	58	7	9	0	5	155
認識率(%)		90.8	91.3	81.9	97.1	95.5	100	97.7	92.4

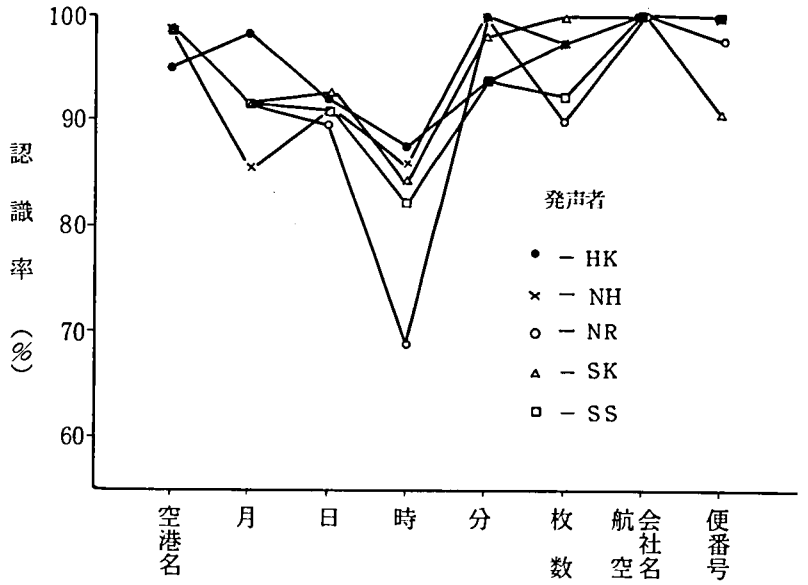


図 9.7　対象別の認識率（実験Ⅰ）

表 9.6　　confusion matrix　　（対象：月）

出 / 入	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	計
1 月							1			1			2
2 月					1								1
3 月													0
4 月							3						3
5 月						1							1
6 月					3								3
7 月	4			2									6
8 月													0
9 月					2								2
10 月				1			1					2	4
11 月													0
12 月													0
												計	22

表 9. 7　Confusion matrix　(対象：日)

出入	1日	2日	3日	4日	5日	6日	7日	8日	9日	10日	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日	31日	計
1日																																0
2日			1																													1
3日			2	1																												3
4日					1																											1
5日	1																															1
6日		1																														0
7日																																0
8日																																0
9日																																0
10日																																0
11日											2										2										1	5
12日															4																	4
13日																																0
14日														1										1								1
15日																									1							2
16日															3										2							5
17日												4																				4
18日																																0
19日															2																	2
20日	2																															2
21日																																0
22日																																0
23日													1																			1
24日																																0
25日																									1							1
26日																									5							5
27日																					3						9					12
28日																						1										1
29日																							9									2
30日																																0
31日																																0
計																																54

表 9. 8　Confusion matrix　(対象：時)

出入	7時	8時	9時	10時	11時	12時	13時	14時	15時	16時	17時	17時	18時	19時	20時	21時	計
7時																	0
8時																	0
9時																	0
10時					3									4			7
11時											2						2
12時				2												1	3
13時																	0
14時				10						1				5			6
15時				2						12				1			5
16時																	0
17時					5											1	6
17時								2									2
18時																	0
19時											5						5
20時				1													1
21時					1												1
計																	58

表 9.9　Confusion matrix　　(対象：分)

出＼入	0分	5分	10分	15分	20分	25分	30分	35分	40分	45分	50分	55分	計
0分													0
5分													0
10分					2								2
15分						1							1
20分			1										1
25分													0
30分													0
35分													0
40分											1		1
45分												1	1
50分									1				1
55分													0
												計	7

表 9.10　Confusion matrix　(対象：便番号)

出＼入	1	2	3	4	5	6	7	8	9	0	便	計
1												0
2												0
3												0
4												0
5												0
6												0
7												0
8												0
9											1	1
0												0
便									4			4
											計	5

実験Ⅱについて空港名の認識結果を表9.11に，空港名以外の単語の認識結果を表9.12に示す。これらを図にしたものが図9.8である。さらに，各単語グループごとのconfusion matrix を

表9.11 認識結果（実験Ⅱ）

対象		空港名
資料の数		670
発声者別の語りの数	SK	3
	NR	0
	HK	1
	NH	0
	SS	1
	KM	3
	NS	3
	TY	1
	SH	0
	IK	0
TOTAL		12
認識率（%）		98.2

表9.12 認識結果（実験Ⅱ）

対象		月	日	時	分	枚数	航空会社	便番号	計
資料の数		240	620	320	240	200	200	220	2040
発声者別の誤りの数	HK	1	2	3	1	3	0	0	10
	NH	2	2	0	0	0	0	0	4
	NR	2	10	1	0	0	0	1	14
	SK	1	8	7	0	1	0	0	17
	SS	4	9	11	1	1	0	0	26
	計	10	31	22	2	5	0	1	71
認識率〔%〕		95.8	95.0	93.1	99.2	97.5	100	99.6	96.5

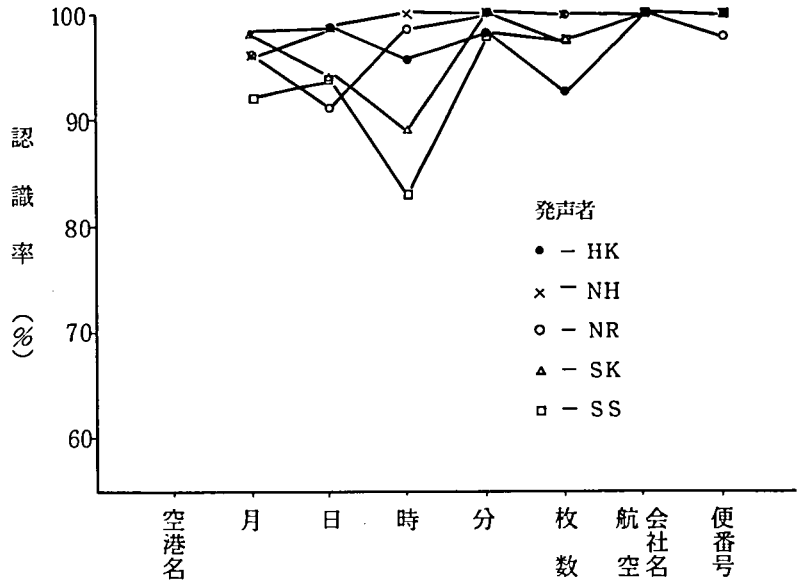


図 9.8　対象別の認識率（実験Ⅱ）

表 9.13 ～表 9.17 に示す。実験Ⅲについては，認識結果を表 9.18 に示す。また，月，日，時の単語グループの confusion matrix を表 9.19 ～表 9.21 に示す。

表 9.13　Confusion matrix　（対象：月）

出／入	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	計
1月							3						3
2月										1			1
3月													0
4月							3						3
5月													0
6月													0
7月	1			2									3
8月													0
9月													0
10月													0
11月													0
12月													0
												計	10

表 9. 14　Confusion matrix　（対象：日）

出＼入	1日	2日	3日	4日	5日	6日	7日	8日	9日	10日	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日	31日	計
1日																																0
2日																																0
3日				2	2																											4
4日																																0
5日																																0
6日																																0
7日																																0
8日																																0
9日																																0
10日																																0
11日																	2															2
12日																																0
13日																																0
14日																										1						1
15日																											1					1
16日															1																	1
17日											1																					1
18日																																0
19日																																0
20日	1																															1
21日											1																					1
22日												1																				1
23日																																0
24日																																0
25日															1																	1
26日																										1			1			2
27日											1						2					5										8
28日																		3														3
29日																			2									1				3
30日																														1		1
31日																																0
																															計	31

表 9. 15　Confusion matrix （対象：時）

出＼入	7時	8時	9時	10時	11時	12時	13時	14時	15時	16時	17時	17時	18時	19時	20時	21時	計
7時																	0
8時																	0
9時																	0
10時																1	1
11時																	0
12時				2	1												3
13時																	0
14時				4													4
15時				2					1		1			4			8
16時																	0
17時					1												1
17時																	0
18時				1													1
19時											1						1
20時				1													1
21時					2												2
																計	22

表 9. 16　Confusion matrix（対象：分）

出 / 入	0分	5分	10分	15分	20分	25分	30分	35分	40分	45分	50分	55分	計
0分													0
5分													0
10分				1									1
15分													0
20分			1										1
25分													0
30分													0
35分													0
40分													0
45分													0
50分													0
55分													0
												計	2

表 9. 17　Confusion matrix（対象：便番号）

出 / 入	1	2	3	4	5	6	7	8	9	0	便	計
1												0
2												0
3												0
4												0
5						1						1
6												0
7												0
8												0
9												0
0												0
便												0
											計	1

表 9.18 認 識 結 果 （実験Ⅲ）

対象＼発声者	NH	IK	SK	SS	FS	KM	NR	KH	WT	HS	SA	KN	計	サンプル数	認識率(%)
1. 月	1	2	0	0	3	3	1	1	2	2	0	3	18	432	95.8
2. 日	3	0	3	3	4	2	0	7	3	2	1	8	36	1116	96.8
3. 時	1	0	6	2	1	1	2	0	3	2	1	3	22	540	95.9
4. 分	0	0	0	1	2	0	0	1	0	1	0	2	7	432	98.4
5. 枚 数	0	1	2	0	0	0	1	0	0	0	1	1	6	360	98.3
6. 航空会社名	0	0	0	0	0	0	0	0	0	0	0	0	0	360	100
7. 空 港 名	3	1	5	6	0	4	1	7	5	3	0	2	37	2412	98.5
8 計	8	4	16	12	10	10	5	16	13	10	3	19	126		
サンプル数	471	471	471	471	471	471	471	471	471	471	471	471		5652	
認識率(%)	98.3	99.2	96.6	97.5	97.9	97.9	98.9	96.6	97.2	97.9	99.4	96.0			97.8

表 9.19 Confusion matrix （対象：月）

出＼入	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	計
1月													0
2月										1			1
3月					1								1
4月										1			1
5月									2				2
6月					1								1
7月	4			5									9
8月													0
9月					3								3
10月													0
11月													0
12月													0
												計	18

表 9. 20　Confusion matrix　(対象：日)

出入	1日	2日	3日	4日	5日	6日	7日	8日	9日	10日	11日	12日	13日	14日	15日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日	31日	計
1日				1																												1
2日				3																												3
3日					1																											1
4日																																0
5日	2																															2
6日																																0
7日																																0
8日																																0
9日																																0
10日								1																								1
11日										1								1				1										3
12日																														1		1
13日																																0
14日			1																													1
15日																																0
16日																2																2
17日											3																					3
18日																																0
19日																														1		1
20日	2																															2
21日											1											1										2
22日												1																				2
23日													4																			1
24日																																4
25日																2																0
26日																	1															2
27日																		1				4										1
28日																																5
29日																																0
30日																																0
31日																																0
計																																36

表 9. 21　Confusion matrix (対象：時)

出入	7時	8時	9時	10時	11時	12時	13時	14時	15時	16時	17時	18時	19時	20時	21時	計
7時																0
8時																0
9時																0
10時				2				1						4		7
11時															2	2
12時																0
13時																0
14時																0
15時					1											1
16時										1						1
17時				4											1	5
18時																0
19時										1						1
20時					3										1	4
21時								1								1
計																22

9.3.4 考　察

まず，実験Ｉの結果について考察を加える。空港名の主な認識誤りを表 9.22 に示す。これを見ると，「高知」(／kochi／)，「隠岐」(／oki／)，「富山」(／toyama／) の誤りが多くなっている。

表 9.22　主な誤り（実験Ｉ）

KOCHI	→	OKI	(11)
KOCHI	→	AOMORI	(2)
OKI	→	OKUSIRI	(9)
OKI	→	KOCHI	(3)
TOYAMA	→	TARAMA	(10)
HIROSIMA	→	KIKAIJIMA	(5)
HIROSIMA	→	YAKUSIMA	(2)
HUKUE	→	UBE	(4)
TOKUNOSIMA	→	TOKUSIMA	(3)
TARAMA	→	KANAZAWA	(3)
NAGOYA	→	KANAZAWA	(2)

「高知」は先頭の子音／k／が検出できなかったことによる。「隠岐」は／k／の後／i／が無声化したためである。「富山」では，拗音／j／(／y／) とその直前／o／のスペクトルの変形が著しいためである。次に，空港名以外の単語の誤りを調べる。これらの対象では，月，日等における数助詞は共通の発声であり，誤りはすべて数字を発声した部分で行っている。代表的な数字の認識誤りを表 9.23 に示す。「10」(／zju／) と「20」(／nizju／) の誤認識は，「20」における／z／のスペクトルパターンが／i／や／j／に似ており，「20」の／izj／と「10」の／j／とが対応づけられやすいこと　「7」(／sichi／) と「1」(／ichi／)，「7」(／sichi／) と 4 (／si／) の発声は人間が聞いても良く似ている。「5」(／go／) と「6」(／roku／) はそれぞれ単独に発声した場合には誤認識は生じないが，時，日などの数助詞が付くと接続部にわたりが生じて誤認識が起こりやすくなる。「9」(／kju／) と「10」(／zju／) は先頭の／k／と／z／の誤認識によるものである。

表 9.23　主 な 誤 り （実験 I）

対象	月	日	時	分	枚 数	計
10 → 20	—	7	6	3	—	16
20 → 10	—	7	2	1	—	10
7 → 1	4	13	6	—	—	23
1 → 7	1	3	2	—	—	6
7 → 4	2	—	—	—	—	2
4 → 7	3	—	—	—	—	3
5 → 6	1	3	2	—	0	16
6 → 5	3	8	0	—	2	13
9 → 10	—	—	0	—	6	6
10 → 9	—	—	0	—	1	1
計	14	41	28	4	9	96

次に実験 II の結果について考察を加える。空港名については実験 I と同じ結果であるが，それ以外の単語については実験 I に比較して誤りは半分になり，全体の認識率が 4.1% 向上している，対象別の confusion matrix を見ると，月，時などの認識率の低かったもの程改善の度合が大きい。各対象に共通した代表的な誤りを表 9.24 に示す。表 9.22 に比較すると「20」から

表 9.24　主 な 誤 り （実験 II）

対象	月	日	時	分	枚 数	計
10 → 20	—	2	1	1	—	4
20 → 10	—	11	3	1	—	15
7 → 1	1	7	1	—	—	9
1 → 7	3	2	0	—	—	5
7 → 4	2	—	—	—	—	2
4 → 7	3	—	—	—	—	3
5 → 6	0	0	1	—	0	1
6 → 5	0	2	0	—	0	2
9 → 10	—	—	0	—	0	0
10 → 9	—	—	0	—	4	4
計	9	24	6	2	4	45

「10」への誤りや「10」から「9」への誤りが増加しているが，「10」から「20」への誤りや「7」から「1」への誤り，「5」と「6」の誤りなどは大幅に減少している。これらの結果から，標準パターンの数を増加した音韻標準パターンⅡが音韻標準パターンⅠより有効であると結論できる。

次に実験Ⅲの結果について考察を加える。実験Ⅱの結果と比較すると分を除いたすべての単語グループで認識率が同等もしくは向上している。confusion matrixの傾向は良く似ており，実験Ⅲでは実験Ⅱと同様に「7」から「1」，「4」への誤り，「10」と「20」の間の誤りが比較的多い。これらの結果から，方法Ⅱが方法Ⅰに比較してより有効であると結論できる。

9.4 あとがき

音韻を単位とした単語音声認識法について述べた。この方法は，標準パターンを音韻単位で蓄えておき，一方，認識対象の単語は音韻系列の形で蓄えておき，これらを併用して入力音声を認識する。このような方法をとることにより，辞書を変えるだけで認識対象語を容易に変更できる，認識対象の単語の一部を発声するだけで標準パターンを登録できる可能性がある，標準パターン，辞書を蓄えておくための記憶容量が少なくてすむ，等の特徴を持つ。

音韻標準パターンとしては2種類を用意した。標準パターンⅠは，音声学的に考えられている音韻に対応した標準パターンを用意するものである。これに対し，標準パターンⅡは，母音間のわたり，調音結合による子音パターンの変動を考慮に入れ，標準パターンⅠに対しいくつかの音韻を追加したものである。

また，DPマッチング法として2種類の方法を提案した。方法Ⅰでは，認識対象の単語を音韻系列で表現し，各音韻には平均的な継続時間の情報を付加してある。このような音韻系列と入力音声とのマッチングを行う。方法Ⅱでは，音韻系列の各音韻に最小，最大の継続時間情報を付加しておき，この条件のもとで入力音声と音韻系列のマッチングを行う。

認識対象として，航空機の座席予約に用いる単語を取り上げ認識実験を行った。その結果，標準パターンについては，標準パターンⅡの方が高い認識性能を持つことがわかった。また，DPマッチング法については，方法Ⅱの方が良い認識結果を示すことがわかった。方法Ⅱは，実時間処理に向いた方法であって，この評価実験により，実時間で動作する認識装置を実現できる見通しが得られた。方法Ⅱに基づいた認識装置の構成については第11章で述べる。

第10章　音韻標準パターンを用いた連続単語音声の認識

10.1　はしがき

第9章では音韻標準パターンを用いた単語音声認識について述べた。本章では，対象を連続単語音声に拡大し，音韻標準パターンを用いた連続単語音声認識について述べる。

第5章では，すでにVCV音節を単位とした連続単語音声認識について述べた。そこで提案した連続単語音声認識法は，すべての単語系列と入力音声とのマッチングをDPを用いて行い，最も類似度の高い単語系列を認識結果とする方法である。この方法は次のような特徴を持っている。

(1)　連続単語を単語単位に区切る，いわゆるセグメンテーションの操作を必要としないため，高い認識率が得られる。

(2)　任意のけた数の連続単語音声の認識が可能である。

さらに，第6章ではDPマッチング法の改良として端点フリーDPマッチング法を提案した。この方法は，2つのパターンを照合する際，必ずしも両パターンの始点，終点を一致させずにマッチングを行う方法であり，一方のパターンが他方のパターンを中に含んだような形になっている場合に有効な方法である。

本章では，これらの方法と第9章で述べた方法を組み合わせ，さらにそれを発展させることにより，3種類の連続単語音声認識方法を提案する，これらの方法は次の特徴を持っている。

(1)　方法Ⅰは入力音声とすべての単語系列のマッチングをDPを用いて行い，最大の類似度を持つ単語系列を認識結果とする方法である。

(2)　方法Ⅱは通常のDPマッチングに加え，音声の終端から時間軸を逆方向にしたDPマッチングを行い，これらを合わせて入力音声と単語系列の最適なマッチングを行う方法である。

(3)　方法Ⅲは入力音声と単語のマッチングを，端点を開放した端点フリーDPマッチングを用いて行うことにより，入力音声と単語との類似度を求めた後，音声の終端から1語ずつ単語を求めて答を得る方法である。

これら3種類の方法を，連続数字音声を用いた認識実験，処理量の比較により比較評価を行う。

10.2 連続単語音声認識方法[142][143]

連続単語音声認識系の構成を図 10.1 に示す。

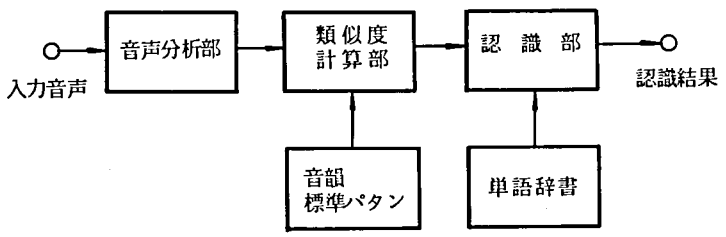


図 10.1　連続単語音声認識系の構成

入力音声はまず音声分析部でAD変換されディジタル音声に変換された後，一定の区間(フレーム）ごとに特徴量が計算され，特徴パラメータの時系列に変換される。次に類似度計算部では，音韻標準パターンと各フレームの特徴パラメータとの類似度が計算される。最後に認識部では，単語辞書に格納されている認識対象の単語と類似度計算部から送られてくる類似度との照合が行われ，認識結果の単語系列が求められる。本章で提案する 3 種類の連続単語音声認識法の違いは認識部の構成の相違によるものであって，他の部分は同一である。以下，各部の処理内容および 3 種類の連続単語音声認識法について述べる。

10.2.1 音声分析

音声分析は第 2 章で述べたのと同じ方法で行われる。入力音声は 3.2 kHz 低域通過フィルタを通った後，標本化周波数 8 kHz の AD 変換器で 11 ビットのディジタル音声に変換される。ディジタル音声は 15 msec ごとのフレームに分けられ，各フレームごとに音声パワーが求められる。あらかじめしきい値を定めておき，音声パワーがしきい値以上になったフレームを音声の始端とする。しきい値以下の音声パワーが一定区間以上続いた時は，最初にしきい値以下になったフレームを音声の終端とする。

音声の分析法としては最尤スペクトル分析法を採用する。最尤スペクトル分析法では，入力音声のスペクトル情報は波形の自己相関関数として求められる。そこで，15 msec のフレームごとに音声波形の自己相関関数を求める。入力音声のフレーム数をN，第iフレームの自己相関関数を

$$\boldsymbol{v} = (v_{i0},\ v_{i1},\ \cdots\cdots,\ v_{ip}) \qquad (10.\ 1)$$

とすると，入力音声は自己相関関数の時系列，

$$V = v_1,\ v_2,\ \cdots\cdots\cdots,\ v_N \qquad (10.\ 2)$$

として表現される。なお，式（10.1）において$v_{i\tau}$ はτ次の自己相関関数を表わす。特にv_{i0}は音声パワーを表わす。

10.2.2 類似度計算

類似度計算部では，入力音声の各フレームの特徴パラメータと音韻標準パターンとの類似度を求める。最尤スペクトル分析法に基づく類似度の尺度としては第2章で提案した尺度を用いる。この尺度を用いる場合，音韻/x/の標準パターンは最尤スペクトルパラメータ

$$Ax = (Ax_0,\ Ax_1,\ \cdots\cdots\cdots,\ Axp) \qquad (10.\ 3)$$

として表現される。また，v_iと/x/の類似度を$\ell(i, x)$とすると，これは次式で定義される。

$$l(i, x) = -\log\left\{\sum_{\tau=0}^{p} A_{x\tau}\ v_{i\tau}\right\} + \log\left\{\sum_{\tau=0}^{p} B_{i\tau}\ v_{i\tau}\right\} \qquad (10.\ 4)$$

ただし，$\{Bi\tau\}$（$\tau = 0, 1, \cdots\cdots\cdots, p$）は$\{v_{i\tau}\}$から求められた最尤スペクトルパラメータである。式（10.4）の右辺の第2項は入力にのみ関係した項であり，入力と各標準パターン間の類似度の相対的な関係のみが問題となるときは式（10.4）右辺の第1項のみを用いれば良く，計算量も少なくてすむ。

類似度計算部の出力結果としては，横軸を入力音声のフレーム番号，縦軸を音韻記号とし，要素を$\ell(i, x)$とした行列が得られる。これを類似度行列と呼ぶ。類似度行列の例を図10.2に示す。

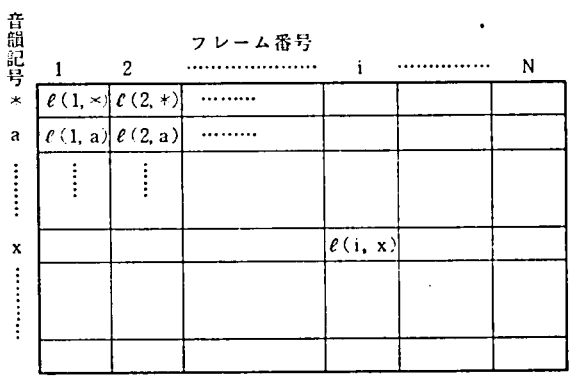


図10.2　類似度行列の例

10.2.3 連続単語音声認識

10.2.3.1 DPマッチング法

認識部では，類似度計算部から送られてくる類似度行列と単語辞書に蓄えられている認識対象の単語との照合を行い，連続単語音声認識を行う。まず，類似度行列と辞書項目との照合に用いられるDPマッチングについて説明する。

単語 w^r の音韻系列表示を

$$w^r = x_1\ x_2\ \cdots\cdots\ x_j\ \cdots\cdots\ x_J \qquad (10.\ 5)$$

$$(r = 1, 2, \cdots\cdots, R)$$

とする。式（10.5）は単語w^rがJ種類の音韻の系列として表現されていることを示している。Rは単語辞書中の音韻系列の数である。各音韻には継続時間の制限がつけられている。x_j の最小継続時間および最大継続時間をそれぞれDj^m，Dj^M（$j = 1, 2, \cdots\cdots, J$）とする。ただし，単位はフレーム数である。単語$w^r$と入力音声$V$とのマッチングは，

$$1 \leqq n_1 < n_2 < \cdots\cdots < n_{J-1} < N \qquad (10.\ 6\ a)$$

$$Dj^m \leqq n_j - n_{j-1} \leqq Dj^M \qquad (10.\ 6\ b)$$

なる制限のもとで次式を求めることによって得られる。

$$L(w^r) = \max_{\{n_j\}} \left\{ \sum_{n=1}^{n_1} \ell(n, x_1) + \sum_{n=n_1+1}^{n_2} \ell(n, x_2) \right.$$

$$\left. + \cdots\cdots + \sum_{n=n_{J-1}+1}^{N} \ell(n, x_J) \right\} \qquad (10.7)$$

ここで$\{n_j\}$は音韻の変化時点を示している。式（10.7）の右辺は，単語w^rの各音韻の最小，最大の継続時間の制限の下で，入力音声とw^rの最大類似度を求めることを意味している。この結果得られた$L(w^r)$が入力音声とw^rの類似度であるとする。

式（10.7）の計算をDPを用いて効率良く求めるには次のようにする。まず，w^rの各音韻をその最大継続時間だけ並べた長さ$\sum_{j=1}^{J} Dj^M$の系列を用意する。次に各音韻記号にフラグを設

け，x_jの最初の$D_j{}^m$個はフラグを1にし，残りの（$D_j{}^M - D_j{}^m$）個には0を与える。w^rを展開してできるこの音韻系列を改めて

$$wa^r = y_1\ y_2 \cdots\cdots y_k \cdots\cdots y_K \qquad (10.8)$$

$$\left(K = \sum_{j=1}^{J} D_j{}^M\right)$$

とする。このとき，v_1，v_2，……，v_iと$y_1\ y_2 \cdots\cdots y_k$の類似度を$L\ (i,\ k)$で表わすと，式（10.7）は次の漸化式表現に書きなおすことができる。

(1) y_kのフラグが1なら

$$L\ (i,\ k) = L\ (i-1,\ k-1) + l\ (i,\ y_k) \qquad (10.9\,a)$$

(2) y_kのフラグが0なら

$$L\ (i,\ k) = \max \begin{cases} L\ (i-1,\ k-1) + l(i,\ y_k) \\ \\ L\ (i,\ k-1) \end{cases} \qquad (10.9\,b)$$

ただし，$L\ (i,\ k)$の初期値は次のようにおく。

$$\left.\begin{aligned} L\ (0,\ 0) &= 0 \\ L\ (i,\ 0) &= L\ (0,\ k) = -\infty \quad (i,\ k \neq 0) \end{aligned}\right\} \qquad (10.10)$$

$L\ (i,\ k)$を順次計算すると，$L\ (w^r) = L\ (N,\ K)$となって，式（10.7）の$L\ (w^r)$が

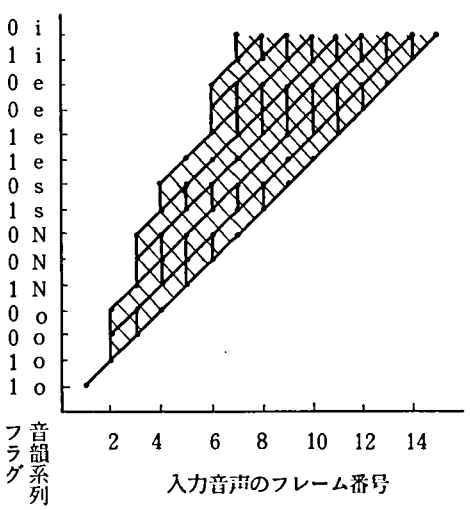


図10.3　入力音声と音韻系列のDPマッチング

求められる。図10.3に入力音声と辞書の音韻系列がマッチングされる様子を示す。図10.3の斜線内部が最適なマッチングを探索する範囲であり，実線で示してあるのが，両者の対応づけが可能なパスを示している。このようなマッチングを行うことにより，入力音声と音韻系列とを時間軸の非常形な伸縮を補正して対応づけることが可能になり，発声するたびに生じる時間長のゆらぎを吸収することができる。

10.2.3.2 連続単語音声認識法Ⅰ

本方法は10.2.3.1で述べたDPマッチング法を拡張して，すべての単語系例と入力音声との，類似度をDPマッチングを用いて求め，それらの中で最も高い類似度を持つ単語系列を認識結果とする方法である。この方法は，第5章で述べた連続単語音声認識法と，第9章で述べた音韻を単位とした単語音声認識法を融合したものである。第5章で述べた方法が，標準パターンとしてVCV音節を用いており学習が容易でないという問題点があったのに対し，本方法では音韻を標準パターンとしており学習が比較的容易であり実用的であるという利点がある。具体的な方法は以下のとおりである。

まず，入力音声の部分系列と単語辞書のマッチングを行う。入力音声の部分系列

$$\boldsymbol{v}_s, \boldsymbol{v}_{s+1}, \cdots\cdots\cdots, \boldsymbol{v}_t \quad (1 \leqq s \leqq t \leqq N) \qquad (10.11)$$

と単語の音韻系列，式（10.5）との類似度を求め，これを

$$LA(s, t \mid r) \qquad (10.12)$$

とする。式（10.12）の値は，10.2.3.1で述べたDPマッチングの手法により求める。式(10.12）をすべての単語について求め，その最大値を

$$LA(s, t) = \max_{r} LA(s, t \mid r) \qquad (10.13)$$

とする。また最大値を与える単語を

$$w(s, t) \qquad (10.14)$$

とする。すべての部分系列について，すなわちすべての s, t $(1 \leqq s \leqq t \leqq N)$ について式（10.13），式（10.14）を求めて記憶しておく。

次に，連続単語音声認識は音声区間全体における類似度和を最大にする単語系列を求めると

いう方針で行う。すなわち，

$$1 \leqq m_1 < m_2 < \cdots\cdots < m_{D-1} < N \qquad (10.15)$$

なる条件のもとで

$$LA = \max_{\{m_j\},D} \Big\{ LA(1, m_1) + LA(m_1+1, m_2) + \\ + LA(m_{D-1}+1, N) \Big\} \qquad (10.16)$$

を満足する単語系列

$$WA = w(1, m_1) \cdot w(m_1+1, m_2) \cdot w(m_{D-1}+1, N) \qquad (10.17)$$

を求めると，これが認識結果の単語系列である。ここで，$\{m_j\}$は単語境界のフレーム番号を示している。式（10.16）をDPを用いて求めるには次のようにする。まず，$LB(t)$を入力音声の部分系列

$$v_1, v_2, \cdots\cdots, v_t \qquad (10.18)$$

とすべての単語系列のマッチングにより得られた最大類似度，また$WB(t)$をその時の単語系列とする。このとき，次の漸化式が成り立つ。

$$LB(t+1) = \max_{0 \leqq i \leqq t} \Big\{ LB(i) + LA(i+1, t+1) \Big\} \qquad (10.19)$$

（ただし，$LB(0) = 0$）

$$WB(t+1) = WB(i^*) \cdot w^* \qquad (10.20)$$

ただし，i^*は式（10.19）を満足するiで，w^*は$LA(i^*+1)$に対応する単語である。式（10.19），式（10.20）に従って順次$LB(t)$，$WB(t)$を求めていくと，

$$LA = LB(N), \qquad WA = WB(N) \qquad (10.21)$$

となる。以上述べた方法による連続単語音声認識の原理を図10.4に示す。式（10.16）からわかるように，連続単語の桁数Dも変数の中に入っている。このことは，あらかじめ桁数がわかっていなくとも，認識結果と同時に桁数も答として得られることを示すものである。すなわち，

この方法の特徴は任意の桁数の連続単語音声が認識できることである。

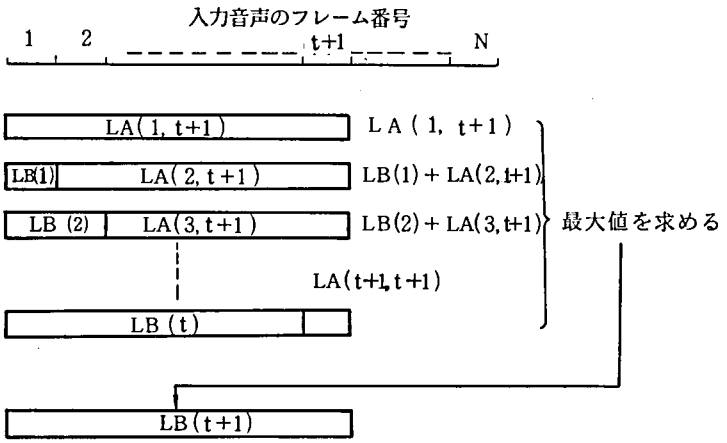


図 10.4.　認識法Ⅰによる連続単語音声認識の原理

10.2.3.3　連続単語音声認識法Ⅱ（逆DPマッチング法）

連続単語音声認識法Ⅰは，すべての単語系列と入力音声とのマッチングを行うという意味で最適な方法であるが，処理量が極めて多くなる欠点を持っている。これは，入力音声のすべての部分系列と単語とのDPマッチングを行う必要があることに起因している。すなわち，異なる s に対して式（10.12）の $L(s, t \mid r)$ を求めるためには，入力音声の各フレームを始点としてDPマッチングを開始する必要があるためDPマッチングの回数が多くなる。これに対し，本項で述べる逆DPマッチングは，必要なDPマッチングの回数を減らすために，導入した方法である。まず，(1)時間軸の正方向へ向うDPマッチングを行い，次に，(2)音声の終端から時間軸の逆方向へ向うDPマッチングを行い，最後に，(3)これら2種類のDPマッチングを統合することにより連続単語音声認識を可能にしたものである。また，本方法では，(4)候補単語の制限によって認識対象の単語の増加に伴う処理量の増加を避けられる特徴も持っている。以下に具体的な方法を示す。

(1)　正方向DPマッチング

入力音声の特徴パラメータの時系列を式（10.2）と同様に

$$\boldsymbol{V} = \boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots\cdots,\ \boldsymbol{v}_N \qquad (10.22)$$

とする。式（10.8）の単語の音韻系列を1ないし2個つないでできる音韻系列を新たに

$$w b^{r} = y_1 \quad y_2 \cdots\cdots y_m \cdots\cdots y_M \qquad (10.23)$$

$$(r = 1,\ 2,\ \cdots\cdots,\ R^2 + R)$$

とする。式（10.22）の先頭から第iフレームまでと式（10.23）の先頭から第M項までの，DPマッチングを10.2.3.1で述べた方法により行う。その結果得られた類似度を

$$LC\ (i,\ M \mid r) \qquad (10.24)$$

とする。入力の各フレームnごとに

$$LC\ (n) = \max_{r} LC\ (n,\ M \mid r) \qquad (10.25)$$

を満足する$LC\ (n)$，およびそれを与える音韻系列に対応する単語系列$WC\ (n)$を求める。

(2) 逆方向マッチング

式（10.22）の特徴パラメータの時系列の時間軸を逆にしたものを

$$\widetilde{V} = v_N,\quad v_{N-1},\ \cdots\cdots,\ v_1 \qquad (10.26)$$

とする。同様に式（10.23）の時間軸を逆にした系列を

$$\widetilde{w b}^{r} = y_M \quad y_{M-1} \cdots\cdots y_1 \qquad (10.27)$$

とする。式（10.26）の先頭から第iフレームまでと式（10.27）の先頭から第M項までのDPマッチングを行い，得られた類似度を

$$\widetilde{LC}\ (i,\ M \mid r) \qquad (10.28)$$

とする。各フレームnごとに

$$\widetilde{LC}\ (n) = \max_{r} \widetilde{LC}\ (n,\ M \mid r) \qquad (10.29)$$

を満足する$\widetilde{LC}\ (n)$および対応する単語系列$\widetilde{WC}\ (n)$を求める。

(3) 正方向DPマッチングと逆方向DPマッチングの統合

最後に，正方向マッチングと逆方向マッチングの結果である式（10.25），式（10.26）を

統合して，

$$\max_{n} \left\{ LC\ (n) + \widetilde{LC}\ (N-n) \right\} \qquad (10.30)$$

を満足するn^*を求めると，単語系列

$$WC\ (n^*)\ \widetilde{WC}\ (N-n^*) \qquad (10.31)$$

が認識結果となる。方法Ⅱの原理を図 10.5 に示す。

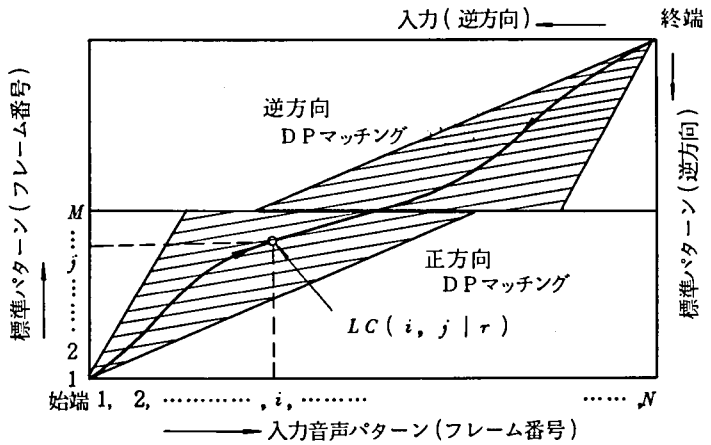


図 10.5　認識法Ⅱ（逆DPマッチング）の原理

以上述べたように，この方式は，4桁以内のすべての単語系列と入力音声とのマッチングを行い，その中で最適のマッチングが行われた結果を認識結果とすることに等しいので，桁数に制限を受けることを除けば方法Ⅰと等価な方法となる。式（10.23），式（10.27）の音韻系列を構成している単語の桁数を増加させるとこの制限を取り除くことができるが，音韻系列の種類が指数関数的に増大するため，4桁程度が実用上の限界であろう。

(4)　候補単語の制限

本方法では，認識対象の単語が増加すると音韻系列の個数（R^2+R）が急激に増加するため好ましくない。そこで，マッチングの回数を減らすため，処理過程で候補単語を上位から数個にしぼる操作を行い，等価的にマッチングの回数を減らす処理を行う。入力音声（式（10. 22））の先頭から第iフレームまでと音韻系列の先頭から第j項までの類似度を$L\ (i, j \mid r)$とすると，あらかじめ定めた入力の第Iフレームにおいて，

$$\max_{j} \quad LC\ (I, j \mid r) \qquad (10.32)$$

を各音韻系列について求める。この値が上位数個のものについてのみDP マッチング計算を続け，それ以外については処理を途中で打切る。これは，図 10.6 に示すように，入力と音韻系列

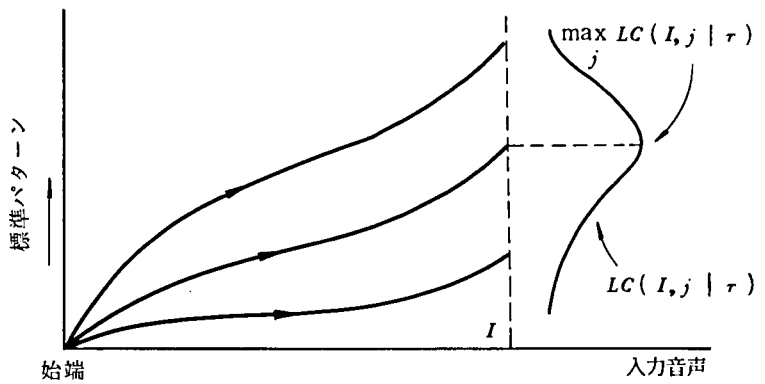


図 10.6　候補単語制限のためのDP マッチング

の始端を一致させ音韻系列の終端を開放した端点フリーDP マッチングを行うことに相当する。この手法を用いて 1 桁目の終端付近で候補を C 個にしぼることにすれば，マッチングすべき音韻系列の個数は（R^2+R）個から（$RC+R$）個へ減少させることができる。以上の説明は正方向マッチングについて行ったが，逆方向マッチングにも同じ処理を全く同様にしてほどこすことができる。この候補単語の制限によって，方法 II における処理量をさらに低減させることが可能である。

10.2.3.4　連続単語音声認識法III

この方法は第 6 章で提案した端点フリーDP マッチング法を連続単語音声認識に適用したものである。端点フリーDP マッチング法は，2 つのパターンを照合する際，両パターンの始端，終端を一致させずにマッチングを行うものである。連続単語音声中には複数個の単語が含まれ，かつそれらの始端，終端がわからないので，音韻系列とのマッチングの際，始端，終端を開放した形でマッチングを行い，最適のマッチングを行って連続単語音声から単語を検出しようとするものである。具体的な手順を以下に示す。

まず，入力音声と音韻系列との端点フリーDP マッチングについて述べる。単語 w^r の音韻系列は方法 I と同様に式（10.5）で表現されるとする。音韻 x_j には平均継続時間長 D_j が対応づけられている。x_j を D_j 個並べることによって作られる音韻系列を改めて，

$$w d^{r} = y_1 \quad y_2 \quad \cdots\cdots\cdots y_k \quad \cdots\cdots\cdots y_K \qquad (10.33)$$

$$\left(K = \sum_{j=1}^{J} Dj\right)$$

と書く。入力音声の第nフレームにおける音韻系列 wd^{r} との類似度を

$$LD(n \mid r) \qquad (10.34)$$

とすると，これは次式で定義される。

$$LD(n \mid r) = \frac{1}{K} \min_{\{\alpha_k\}} \sum_{k=1}^{K} \ell(\alpha_k, y_k) \qquad (10.35)$$

ここで，$\{\ \alpha_k\ \}$はkの関数であり，

$$\alpha_1 \leqq \alpha_2 \leqq \cdots\cdots\cdots \leqq \alpha_K = n \qquad (10.36)$$

なる関係があると同時に，$\ell(\alpha_k, y_k)$は図10.7に示したパスにそって加えてゆく必要がある。

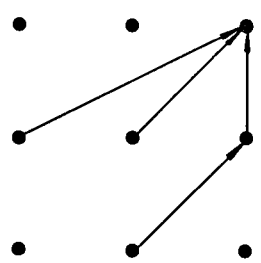


図10.7　方法Ⅲにおける DP パス

このとき，式（10.35）を漸化式表現すると次のようになる。

$$LD(i, k \mid r)$$

$$= \max \begin{cases} LD(i-2, k-1 \mid r) + \ell(i, y_k) \\ LD(i-1, k-1 \mid r) + \ell(i, y_k) \\ LD(i-1, k-2 \mid r) + \ell(i, y_{k-1}) + \ell(i, y_k) \end{cases} \qquad (10.37)$$

$$LD(n \mid r) = \frac{1}{K}\ LD(n, K \mid r) \qquad (10.38)$$

ただし，初期値は次のようにおく。

$$\left.\begin{aligned} LD\ (i,-1\mid r)=LD\ (0,k\mid r)=LD\ (-1,k\mid r)=-\infty \\ LD\ (i,\ 1\mid r)=\ell(i,\ y_1) \end{aligned}\right\} \qquad (10.39)$$

式（10.36）のα_1はマッチングの始端であり，これが変数であることから，始端を開放したマッチングを行うことになる。しかも，点（i, k）へ至るパスの長さはいずれもkであるから，式（10.37）で種々の異なる3点から出発したDPパスを比較する際，公平な比較が行われる。このDPマッチングの様子を図10.8に示す。

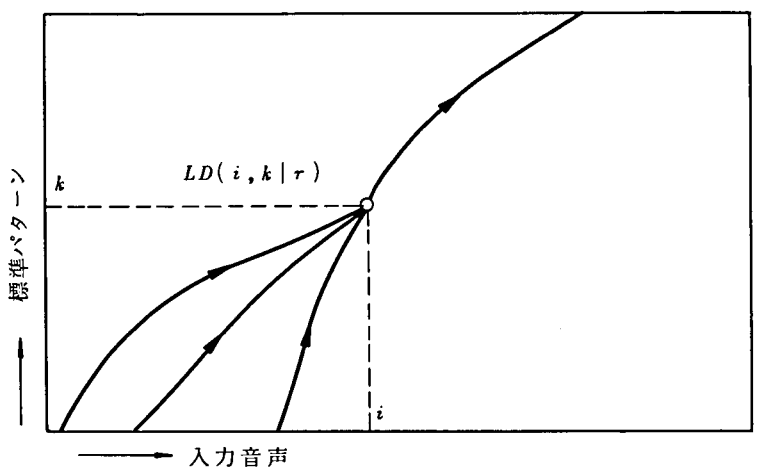


図10.8　方法ⅢにおけるDPマッチング

入力音声の第nフレームを終端とするDPマッチングパスの始端を知るために関数$T(n\mid r)$を導入する。これは次の漸化式で求めることができる。

$$T\ (i,\ 1\mid r)=i \qquad (10.40)$$

$$T\ (i,\ k\mid r)\begin{cases} T\ (i-2,\ k-1\mid r)\text{（式(10.37)の第1式が選ばれた場合）} \\ T\ (i-1,\ k-1\mid r)\text{（式(10.37)の第2式が選ばれた場合）} \\ T\ (i-1,\ k-2\mid r)\text{（式(10.37)の第3式が選ばれた場合）} \end{cases} \qquad (10.41)$$

$$T\ (n\mid r)=T\ (n,\ k\mid r) \qquad (10.42)$$

すべての音韻系列について，$\{LD(n\mid r)\}$，$\{T(n\mid r)\}$を求めておく。

次に，連続単語音声認識は次の手順で行う。

(1)　入力音声の終端Nにおいて最大の類似度$\max LD(N \mid r)$を与える単語wd_1，およびその始端n_1を求める。

(2)　前段の処理で決定した単語の始端をn_iとするとき

$$n_i - T_1 \leqq n \leqq n_i + T_1 \quad (T_1 \text{はあらかじめ定めたしきい値}) \qquad (10.43)$$

なる時間において各$LD(n \mid r)$の最大値を求める。単語ごとに求めた最大値の中で，最大の値を与える単語をwd_{i+1}とする。また，その始端をn_{i+1}とする。

(3)

$$n_{i+1} > T_2 \quad (T_2 \text{はしきい値}) \qquad (10.44)$$

なら(2)の処理をくり返すことによってさらに単語を求める。

$$n_{i+1} \leqq T_2 \qquad (10.45)$$

なら，単語系列が求まったとして処理を終了する。

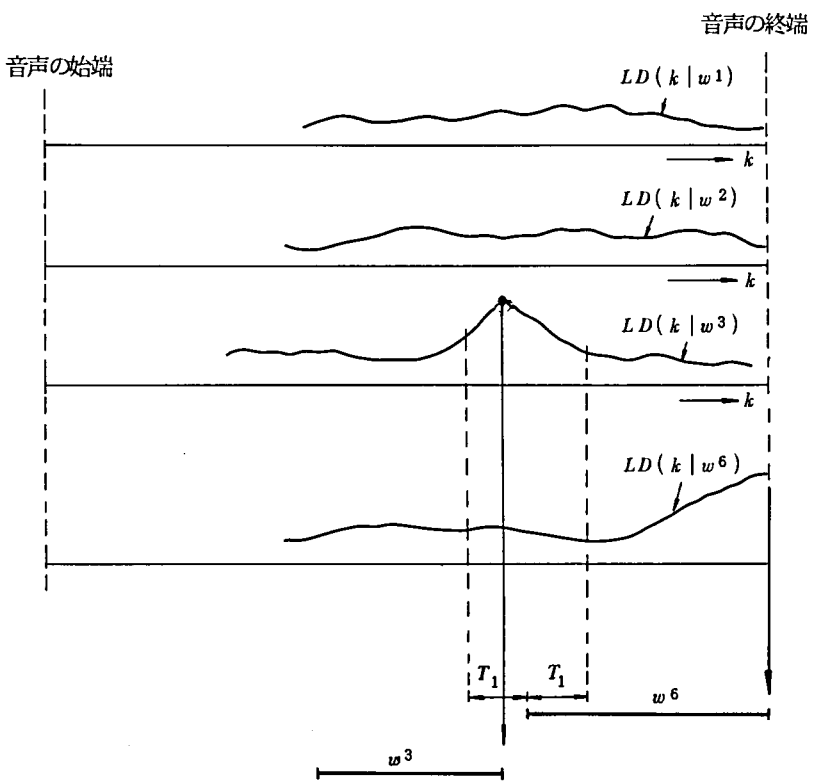


図10.9　方法Ⅲによる連続単語認識の手順

(4) 以上のように，入力音声の語尾から順次単語を決定し，得られた単語系列

$$wd_I \; wd_{I-1} \cdots\cdots\cdots\cdots wd_1 \qquad (10.46)$$

を認識結果とする。

以上の手順によって単語系列を求めて行く様子を図10.9に示す。

10.3 連続単語音声認識方法の評価[142][143]

10.3.1 認識実験による評価

10.2.3で述べた連続単語音声認識法の評価を行うために，計算機シミュレーションによる認識実験を行った。

10.3.1.1 認識対象

連続単語音声の応用範囲として最も多いのは，数字データを連続的に計算機などへ入力する場合であろう。そこで，数字音声を認識対象として選んだ。認識対象の各数字，およびその読みを表10.1に示す。

表10.1 認識対象

単　語	読　み
1	イチ
2	ニ
3	サン
4	ヨン
5	ゴ
6	ロク
7	ナナ
8	ハチ
9	キュー
0	レー

10.3.1.2 音韻標準パターン

音韻標準パターンは，表 10.2 に示した 21 種類のものを用いた。これらのうち，/i′/,/u′/は

表 10.2 音韻標準パタンの種類

母音	a, i, u, e, o, i′, u′, $\tilde{i}$,
母音のわたり	i o, i u
子音	g, r, n_i, N, h, ch, k_i, k_u, s, n
無音区間	*

それぞれ無声化した/i/,/u/を表す。また，/$\tilde{i}$/,/n_i/は，“2”における鼻音化した/i/と，後続の母音の影響を受けて変形した/n/を表す。また，/k/は後続の母音の影響を受けて変形しやすいので，母音/i/,/u/に対応してそれぞれ/k_i/,/k_u/を用意した。/*/は，無音部分に対応する標準パターンで，語中のポーズを表現するのに用いる。

10.3.1.3 単語辞書

表 10.3 に単語辞書の内容を示す。実際には，10.2.3.1 で述べたように各音韻には最小，最

表 10.3 単語辞書の内容

数字	音韻系列
1	* i * ch i * i * ch i′ * i * ch
2	* n_i $\tilde{i}$
3	* s a N
4	* i io o N
5	* g o
6	* r o * k_u u * r o * k_u u′ * r o * k_u
7	* n a n a
8	* h a * ch i * h a * ch i′ * h a * ch
9	* ki i iu u
0	* r e i

大の継続時間の制限がついているが，ここでは省略した。“1”，“2”，“8”，に対しては，語尾の母音が無声化しやすいので，それを考慮して複数個の音韻系列を用意した。各音韻系列の先頭には無音区間を表す/*/がついているが，これは，連続数字の各桁間に短いポーズが入る場合を考慮したためである。

10.3.1.4 認識結果

1～4桁の連続数字を認識対象とする。この範囲内では，方法Ⅰと方法Ⅱ（逆DPマッチング）は候補単語の制限の操作を除いては方式的に等価である。そこでまず，候補単語の制限によって認識率の低下が生じないかどうか調べる予備実験を行った。2名の男性発声者が1～4桁の連続単語を発声した計400サンプルを用いた。方法Ⅱで入力音声の第15フレーム（式（10．32）において$I=15$の場合）において候補単語を4個にしぼる操作を行うと，正しい候補が含まれる割合は100％であった。この結果から，候補単語の制限による認識率の低下はほとんどなく，方法Ⅰと方法Ⅱは認識性能に関しては同等と考えてよいことになる。従って，以下の実験は，方法Ⅰ，方法Ⅲで行った。

11名の男性発声者が1～4桁の連続数字を静かな室で発声した2,200サンプルの音声データをテストサンプルとして用いた。これらのサンプル中に含まれる数字は計5,500個である。音韻標準パターンは各発声者が別に発声した10数字の音声データを用いて学習により各発声者ごとに作成した。方法Ⅰを用いた場合の認識結果を表10.4に示す。誤りの数および認識率は数字単位で算出してある。平均の認識率は99.7％である。これは，単独に発声した数字音声の認識

表10.4　連続数字音声の認識結果（方法Ⅰ）

	誤り の 数											認識率
発声者	KI	SF	KK	KH	RN	SI	HN	SS	SA	NA	NK	(%)
1桁数字	0	0	0	0	0	0	0	0	1	0	0	99.8
2桁数字	0	2	0	0	0	0	0	0	0	0	1	99.7
3桁数字	0	1	0	0	1	0	0	0	0	0	0	99.9
4桁数字	1	4	0	1	0	1	0	1	0	0	3	99.5
認識率(%)	99.8	98.6	100	99.8	99.8	99.8	100	99.8	99.8	100	99.2	99.7

結果[19]とほとんど同じであり，連続数字にもかかわらず高い認識率が得られることを示している。また，いずれの桁数の数字に対しても高い認識率が得られている。表10.5には認識誤りの

表 10.5　認識誤りのリスト

K I	8689→8889	K H	4467→5467
S F	28→ 08	R N	363→ 343
	39→ 89	S I	7533→3537
	233→ 237	S S	1686→19586
	1769→1789	N K	35→ 76
	5680→ 580		1609→ 909
	8518→8418		3209→3299
	1126→ 116		
S A	6→ 5		

リストを示す。異なった桁数への誤りは誤り全体の約 1／4で比較的少ない。これは，挿入誤り，脱落誤りが少ないことを示している。同じ桁数間の誤りでは，“6”と“8”の間の誤り，“3”と“7”の間の誤りのように単独発声の数字場合と同じ誤りの傾向がでている。

次に，方法Ⅲを用いた場合の認識結果を表 10.6 に示す。平均の認識率は 98.1 %で，方法 I

表 10.6　連続数字音声の認識結果（方法Ⅲ）

	誤　り　の　数											認識率
発声者	KI	SF	KK	KH	RN	SI	HN	SS	SA	NA	NK	(%)
1 桁数字	0	0	0	0	0	0	0	0	1	0	0	99.8
2 桁数字	1	1	3	1	1	3	1	1	0	1	1	98.7
3 桁数字	0	1	5	1	9	0	2	0	3	6	1	98.3
4 桁数字	8	5	3	4	10	6	5	3	5	2	11	97.2
認識率(%)	98.2	98.6	97.8	98.8	96.0	98.2	98.4	99.2	98.2	98.2	97.4	98.1

を用いた場合に比較して 1.6 %低下している。また，桁数が多くなると共に，認識率が低下する傾向がみられる。これは，方法 I が音声区間全体と単語系列とのマッチングを行い最適の単語系列を求めているのに対し，方法Ⅲでは連続単語中の単語を 1 桁ずつ順に求める方法をとっているためである。方法Ⅲの認識誤りを分析すると次の誤りが多い。

(1)　“1”の挿入誤りと脱落誤り。

(2)　“3”，“4”の認識誤り。

(3) "6" の認識誤りと脱落誤り。

挿入誤りと脱落誤りに対しては，数字の継続時間情報，特に，連続数字音声中の各桁の数字の長さがほぼ等しくなることを積極的に利用すれば，誤りを減らすことができよう。

10.3.2 処理量による評価

方法Ⅰ～Ⅲの処理量の比較を表10.7に示す。いずれの方法でも類似度の計算までは同一の処

表10.7 処理量の比較

方法	処理量	$N=100$, $R=20$, $C=4$
方法 Ⅰ	NR	2000
方法 Ⅱ	$2(R^2+R)$	840
方法 Ⅱ（候補単語制限）	$2(CR+R)$	200
方法 Ⅲ	R	20

理であるため，必要なDPマッチングの回数で処理量の比較を行っている。方法Ⅰでは入力音声の各フレームごとにDPマッチングを開始する必要があるので，DPマッチングの回数は$N\times R$（N；入力音声のフレーム数，R；認識対象の単語数）である。これに対し，方法Ⅱでは，マッチングすべき音韻系列の数はR^2+Rであり，正方向，逆方向の2種類のDPマッチングを行うから，必要なDPマッチングの回数は$2\times(R^2+R)$である。さらに，候補単語の切りすてを行って音韻系列の1桁目の終りで候補をC個にしぼるとDPマッチングの回数は$2(CR+R)$になる。また，方法Ⅲでは始端を開放したDPマッチングを行っており，各音韻系列に対してそれぞれ1回のDPマッチングでよいから，回数はRでよい。表9.7には一例として，$N=100$, $R=20$, $C=4$，とした場合の具体的な数値もあわせて示した。この場合，候補単語の制限を行うと方法Ⅱは方法Ⅰの1／10の処理量でよい。同様に方法Ⅲは方法Ⅰの1／100の処理量でよい。

10.4 あとがき

連続単語音声を認識する3種類の新らしい方法を提案した。これらの方法はいずれも音韻単位の標準パターンと音韻系列で表現された単語辞書を用いている。各方法の方式上の特徴は次のような点にある。

(1) 方法Ⅰは，DPを用いてすべての単語系列と入力音声との類似度を求め，最大の類似度を持つ単語系列を認識結果とする方法である。

(2) 方法Ⅱは，音声の始端から入力音声と音韻系列とのマッチングを行う通常のDPマッチングに加え，音声の終端から時間軸の逆方向にDPマッチングを行い，これら2種類のマッチングを統合して入力音声と単語系列の類似度を求め，最大の類似度を持つ単語系列を認識結果とする方法である。

(3) 方法Ⅲでは，まず入力音声と音韻系列のマッチングを入力音声の始端，終端を開放した条件で行うことによって，入力音声と音韻系列との類似度を求める。次に，この結果を利用して入力音声の終端から1語ずつ認識結果を求め，答の単語系列を得る。

男性発声者を用いた1～4桁の連続数字の認識実験，および処理量の比較により，これらの方法の性能上の特徴は次の点にあることがわかった。

(1) 方法Ⅰは任意の桁数の連続単語音声が認識できる利点がある。また，連続数字の認識実験では99.7％という高い認識率が得られた。

(2) 方法Ⅱは認識対象の単語数，連続単語の桁数が限定される欠点があるが，この範囲では方法Ⅰとほぼ等価であり，高い認識率が得られる。また，DPマッチングに要する処理量が方法Ⅰに比較して1／10程度ですむ利点がある。

(3) 方法Ⅲは方法Ⅰと同様に桁数の制限を受けない。また，DPマッチングに要する処理量は3種類の方法の中で最も少ない。ただし，認識性能は方法Ⅰに比較して少し劣る。

連続単語音声認識装置を作る場合には，認識対象の語彙数，連続単語の桁数，必要な認識性能などから最適な方法を選択すればよい。連続単語音声認識装置については次章で述べる。

第11章　日本語音声認識システムのハードウェア化

11.1　はしがき

前章まで日本語音声を対象とした音声認識システムについて述べてきた。システム構成の基本原理として，VCV音節を標準パターンにとる方式と，音韻を標準パターンにとる方式の2つについて検討し，それらの有効性について示した。これらのシステムはいずれも計算機シミュレーションに基づいたものであり，認識性能の追求を第1の目的としており，処理時間等については特に注意をはらわなかった。しかしながら，音声認識の主たる目的が実用に供しうる認識装置の実現である以上，提案した方式の有効性を最終的に示すには，認識システムの高速化，ハードウェア化を計り，実時間もしくはそれに近い処理時間で動作する認識システム，認識装置を作成する必要がある。

本章では，以上のような観点から著者が試作した音声認識装置およびオンライン音声認識システムについて述べる。まず，音韻を標準パターンとして用いた単語音声認識装置，連続単語音声認識装置の構成と性能評価実験について述べる。次に，計算機とハードウェアの組合わせによるオンライン会話音声認識システムについて述べる。

11.2　単語音声認識装置[140]

第9章では音韻を単位とした単語音声認識方法について述べた。この方法は音韻単位の標準パターンと音韻系列で記述した単語辞書を併用する方法であって，(1)認識対象の単語が多い場合でも標準パターンの話者への適合が簡素化できる可能性がある[52][53]，(2)単語辞書の入れ替えによって認識対象の変更が容易に行える，(3)標準パターンと単語辞書を蓄えるためのメモリが少なくてすむ，などの利点を持っており多数語彙の認識に適した方式である。この方式に基づいた単語音声認識におけるDPマッチングの方法として第9章では2つの方法を提案し，評価実験を行った。その結果，方法Ⅱが認識性能に優れており，かつ，実時間処理に適していることが明らかになった。そこでここでは，第9章で述べた方式Ⅱを採用することとし，それに基づいて試作した単語音声認識装置の構成と性能について述べる。

11.2.1　認識装置の構成

本装置は，相関分析に基づいた特徴抽出，音韻標準パターンとの類似度計算，単語の音韻系列との類似度を求めるDP計算のように比較的単純で処理量の多い演算が多いため，主な処理は専用のハードウェアによって行い，処理全体の制御のためにマイクロプロセッサを用いるという構成をとった。

装置の構成を図11.1に示す。またその外観を図11.2に示す。本装置は，入力音声の特徴量

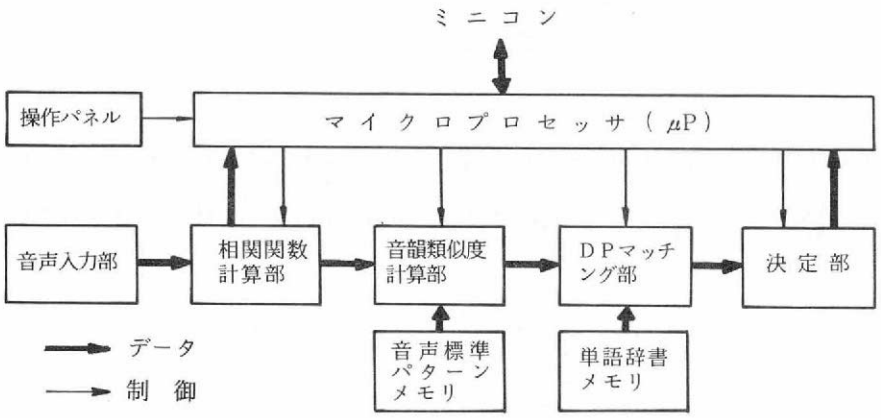


図11.1　単語音声認識装置の構成

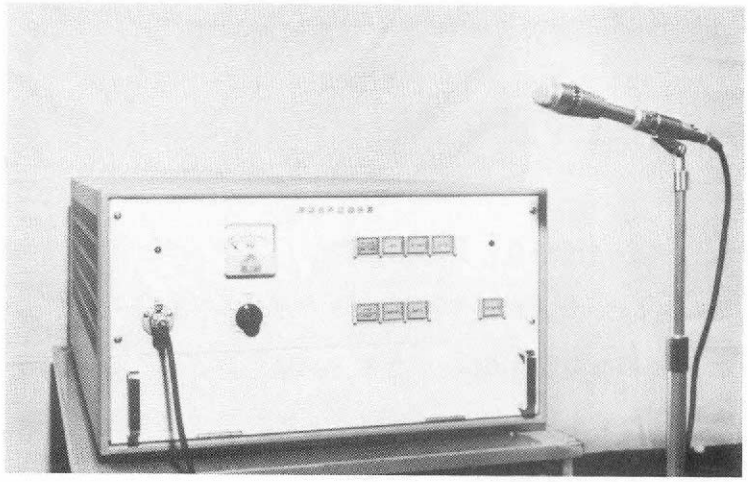


図11.2　単語音声認識装置の外観

として一定長のフレーム毎に自己相関関数を計算する相関関数計算部，標準パターンメモリに蓄えられた音韻の標準パターンとの類似度を計算する類似度計算部，単語辞書メモリに蓄えら

れた音韻系列とマッチングを行って，単語との類似度を計算する単語DPマッチング部，類似度の最も大きい音韻系列の単語を選び出して認識結果とする決定部，音声区間の検出と全体の制御を行うマイクロプロセッサ，認識結果を表示するディスプレイ装置などから構成されている。各部はそれぞれ独立したハードウェアで作られているため，音声の入力と並行して演算が進む，いわゆる実時間処理が可能となっている。この認識装置は単体で動作すると同時にミニコン（NEAC 3200／70）とのインターフェースが内部に組み込まれており，プログラムによる認識動作の開始や，DMAチャネルを用いた標準パターンデータ，単語辞書データ，認識結果，マイクロプロセッサの制御プログラムなどの転送も行うことができる。

11.2.2 各部の動作

11.2.2.1 音声入力部

マイクロホンから入力される音声に対して，6 dB／oct の高域強調を行い，しゃ断周波数3.2 kHz の低域通過フイルタを通した後に，サンプリング周波数 8 kHz，量子化精度 12ビット（符号＋11ビット）でAD変換を行う。

11.2.2.2 相関関数計算部

15 msec のフレーム周期で，10次までの自己相関関数を計算する。標本値を $\{ y_t \}$ とすると，自己相関関数は次式で計算される。

$$v_\tau = \sum_{t=1}^{120-\tau} y_t \cdot y_{t+\tau} \quad (0 \leqq \tau \leqq 10) \qquad (11.1)$$

この相関関数は 32 ビットで計算された後，v_0 のMSBが 1 になるまで左シフトされる。次に他の v_τ も同じビット数だけ左シフトされる。その後各 v_τ（$\tau \geqq 0$）の 32 ビットのうち上位 16ビットを取り出す。この操作をシフト正規化と呼び精度の低下をおさえながらビット数を削減できる。この様子を図 11.3 に示す。シフト後の相関関数を $\{ u_\tau \}$ とすると

$$\left.\begin{aligned} &0.5 \leqq u_0 \leqq 1 \\ &|u_\tau| < 1 \qquad (\tau \geqq 1) \end{aligned}\right\} \qquad (11.2)$$

である。

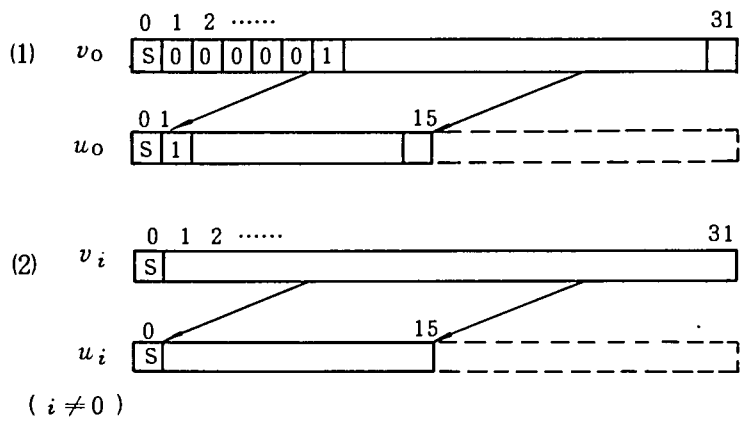


図 11.3　シフト正規化

(1)　v_0 のMSBが1になるまで左シフトした後上位16ビットをとり出す。

(2)　他の v_i についても同じビット数だけ左シフトした後上位16ビットをとり出す。

11.2.2.3　音韻類似度計算部

フレーム毎に，各音韻の標準パターンに対する類似度を計算する。標準パターンを表わす最尤スペクトルパラメータを $\{A_\tau\}_{\tau=0}^{10}$ とすると類似度は次式で定義される。

$$\ell = -\log \sum_{\tau=0}^{10} A_\tau \frac{u_\tau}{u_0} \qquad (11.3)$$

これは，第2章で論じた相対的な類似度である式（2.34）において，相関関数のかわりに，相関係数を用いたものである。相関係数を用いても，入力にのみ関係する値が付加されるだけであり，式（11.3）は式（2.34）と同じ意味を持っている。

式（11.3）の計算を固定小数点演算で行うため，$\{A_\tau\}$ を正規化しておく。$\{A_\tau\}$ のうち，A_0 は第2章で述べたように線形予測係数の相関関数の0次の値，A_τ（$\tau \leq 1$）は τ 次の2倍の値として定義されているので，次のような正規化を行う。

$$a_\tau = \frac{A_\tau}{2A_0} \qquad (11.4)$$

これによって

$$a_0 = 0.5, \quad |a_\tau| < 1 \quad (\tau \geq 1) \qquad (11.5)$$

となる。このとき，式（11.3）は次のように書きなおすことができる。

$$\ell = -\log \sum_{\tau=0}^{10} a_\tau u_\tau + \log u_0 - \log 2A_0 \qquad (11.6)$$

したがって，標準パターンとしては，$\{A_\tau\}$の代りに$\{a_\tau\}_{\tau=1}^{10}$と$-\log 2A_0$を用意しておけばよい。式（11.6）の対数計算は，対数値が書き込まれたテーブルを参照することにより行う。

この類似度計算部は，最大63個の標準パターンに対する類似度が計算できるように設計されている。さらに，音声パワー情報を有効に利用するため，音声パワー値が閾値以上のフレームでは雑音の標準パターンとの類似度を低い値に設定するようになっている。これによって，“イチ”，“ロク”のように語中に無音区間を含む単語の音韻系列が，“ニ”，“ゴ”のようなそうでない単語に誤って対応づけられることを防いでいる。

11.2.2.4 DPマッチング部

単語辞書中には認識対象の単語が音韻系列の形で蓄えられている。DPマッチング部では入力音声と音韻系列とのDPマッチングを行う。DPマッチングの方法は，第2章で述べた方法Ⅱと同じであるが，実時間処理という観点から再び説明する。

従来のDPマッチング法[150]では，単語辞書に蓄える音韻系列の各音韻に平均的な継続時間を与えておき，DPを行うマトリクス上で入力音声の始端と終端を結ぶ対角線上に一定幅の整合窓をかけてマッチングを行っていた。しかしこの方法は，入力音声の終端が定まらないと整合窓がかけられず，マッチングを開始できないという欠点があった。本装置で採用している方法は，整合窓の代りに音韻系列の各音韻に最大，最小の継続時間を与え，入力音声のフレーム周期に同期してマッチングが行える実時間処理向のDPマッチング法である。以下にその方法を述べる。

入力音声の長さをNフレーム，その音声とマッチングを行う単語の音韻系列による表示をx_1 x_2 …… x_j …… x_J とし，各音韻に最小継続時間D_j^mと最大継続時間D_j^Mの制限がつけられているとする，入力音声とこの音韻系列とのマッチングでは

$$\left.\begin{aligned} &1 \leqq n_1 \leqq n_2 \cdots\cdots \leqq n_{J-1} \leqq N \\ &D_j^m \leqq n_j - n_{j-1} \leqq D_j^M \quad (j = 1, 2, \cdots\cdots, J) \end{aligned}\right\} \qquad (11.7)$$

（ただし，$n_0 = 1$，$n_J = N$）

なる条件のもとで

$$L = \max_{\{n_j\}} \left\{ \sum_{n=1}^{n_1} \ell(n, x_1) + \sum_{n=n_1+1}^{n_2} \ell(n, x_2) \right.$$
$$\left. \cdots\cdots + \sum_{n=n_{J-1}+1}^{N} \ell(n, x_J) \right\} \qquad (11.8)$$

を満足する$J-1$個の音韻の変化時点$\{n_j\}$を求める。ここで$\ell(n, x)$はnフレーム目の音声と音韻/x/との類似度である。式（11.8）は，各音韻の最小，最大の継続時間制限のもとで，入力と音韻系列の最適の対応づけを求めることを意味している。本装置では，この計算をDPを用いて次のように効率よく計算している。

単語辞書に登録する音韻系列として，各音韻の最大継続時間だけ同じ音韻記号を並べた長さ$\sum_{j=1}^{J} D_j^M$の系列を用意する。各音韻記号に1ビットのフラグを設けてD_j^M個の音韻記号のうちはじめのD_j^m個のフラグを1に，残りを0にしておく。入力音声の1～iフレームと，音韻系列の1～jフレームの類似度を$L(i, j)$とすると，次の漸化式が成り立つ。

(1) j番目の音韻記号のフラグが1なら

$$L(i, j) = L(i-1, j-1) + \ell(i, x) \qquad (11.9\,a)$$

(2) j番目の音韻記号のフラグが0なら

$$L(i, j) = \max \left\{ \begin{array}{l} L(i-1, j-1) + \ell(i, x) \\ \\ L(i, j-1) \end{array} \right.$$

この漸化式を次の初期条件のもとで$1 \leqq i \leqq N$，$1 \leqq j \leqq M$（$M = \sum_{j=1}^{J} D_j^M$）の範囲で計算する。

$$\left. \begin{array}{l} L(0, 0) = 0 \\ \\ L(0, j) = L(i, 0) = -\infty \\ \\ (i = 1, 2, \cdots\cdots, N\ ;\ j = 1, 2, \cdots\cdots, M) \end{array} \right\} \qquad (11.10\,b)$$

この結果，式（11.8）のLは次式で与えられる。

$$L = L(N, M) \qquad (11.11)$$

図11.4に入力音声と音韻系列のマッチングの様子を示す。実際には，式（11.9 a），(11.9 b)

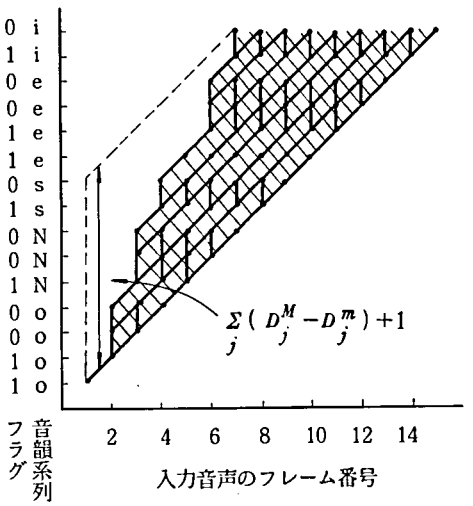


図 11.4　入力音声と音韻系列のDPマッチング

の漸化式は $i \leqq j \leqq i + \sum_{j=1}^{J} (D_j^M - D_j^m)$ の範囲で計算すればよいので，音韻系列毎に

$$\sum_{j=1}^{J} (D_j^M - D_j^m) + 1 \qquad (11.12)$$

の個数の途中結果を保持しておくワークメモリを用意して，その内容を入力音声のフレーム毎に更新すればよい。以上の処理により実時間処理を実現している。

11.2.2.5　決定部

決定部では，DPマッチング部で計算された入力音声と辞書中の各音韻系列との類似度和の比較を行い，大きい順に第1位から第3位まで選んで出力する。

11.2.2.6　マイクロプロセッサによる制御

マイクロプロセッサは，音声の始端と終端の検出，DPマッチング部と単語決定部の制御を行う。すなわち，相関関数計算部から1フレーム毎に音声パワー（相関関数の0次の値）を受けとり，それがあらかじめ定めた閾値をこえたかどうかで始端の検出を行う。始端が検出されるとワークメモリに初期値を設定してDPマッチングを開始させる。さらにフレーム毎にパワーを調べ続け，パワーが閾値よりも小さくなったとき，その直前のフレームを音声の終端の候補とする。そして単語決定部より第1位から第3位までの単語番号と類似度和を読みとる。その後パワーが閾値よりも小さい区間が一定時間以上続いたならば，候補の単語を認識結果としてディスプレイに表示する。一定時間以内に再びパワーの値が大きくなった場合は，単語中の無

音区間とみなして先の結果をキャンセルし，さらにマッチングを続け，終端の候補を探す。全体のタイムチャートを図11.5に示す。

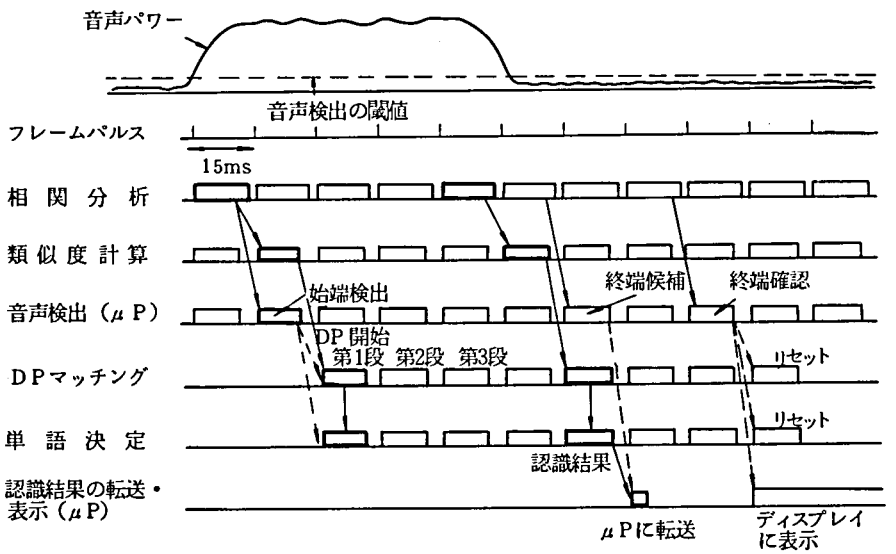


図11.5　認識動作のタイムチャート

注）μpはマイクロプロセッサの略

11.2.3　認識装置の仕様

本装置の仕様を表11.1に示す。またその記憶容量のみを取り出して表11.2に示す。標準パターンメモリには最大63個の標準パターンを蓄えることができる。単語辞書とワークメモリに対しては，音韻系列の先頭アドレスと最終アドレスを示すアドレステーブルが別に用意されており，音韻系列や演算の途中結果をつめて蓄えるようにしている。音韻系列の各音韻を表わすためにフラグと合わせて8ビットずつ使われているので，単語辞書メモリは合計16 kの長さの音韻系列を蓄えることができる。また類似度和は24ビットで表現されており，ワークメモリは合計8 kの途中結果を蓄えることができる。このように，単語辞書メモリ，ワークメモリは不等長で使われるため，認識できる単語数は音韻系列の与え方によって異なる。後で述べる空港名のように1つの音韻系列の平均長さを82，必要なワークメモリを60とすると，マッチングできる音韻系列の上限の数は，ワークメモリの数で決まり，約130個となる。ワークメモリは1フレーム（15 msec）内に書き換えができる数まで増設できるので，単語数を増やすことは可能である。なお，アドレステーブルには最大255個までの音韻系列のアドレスを蓄えられるようになっている。図11.6は本認識法をもとにして，単語数と記憶容量の関係を調べたもの

表 11.1　単語音声認識装置の仕様

サ　イ　ズ	500 W × 300 H × 800 L　　(mm)	
音　声　入　力　部	等化器	6 dB ／ oct の高域強調
	L．P．F．	f_0 = 3.2 kHz，4 kHz
	AD変換器	8 kHz サンプリング 12 ビット量子化
相関関数計算部	フレーム周期	15 ms
	分析次数	0次～10次
	相関関数の表現形式	v_0 のMSBが1になるようシフト正規化，16 ビット
音韻類似度計算部	標準パタンの表現形式	最尤スペクトルパラメータを±1の範囲に正規化，16 ビット
	標準パターン数	最大63 個
	類似度の計算方法	最尤スペクトル分析法に基づく類似度
DP マッチング部	単語辞書の表現形式	各音韻の最小，最大継続時間を指定した音韻系列
	単語辞書の音韻系列数	最大255 個
	DPマッチング法	パイプライン方式による完全時間処理，24 ビット
単　語　決　定　部	認識結果	類似度和第1位～第3位の単語を選択
マイクロプロセッサ（μP）		M 6800

である。点線で示したのは，標準パターンを単語単位で蓄えた場合の単語数と記憶容量の関係であり，この場合に比べてメモリの増加の割合は 1／4以下である。

表 11.2　単語音声認識装置の記憶容量

内　　容	容　量
標準パターンメモリ	2kB
単語辞書メモリ	16kB
ワークメモリ	24kB
マイクロプロセッサ用メモリ	4kB
その他（表示データ／対数値テーブル／アドレステーブル）	9kB
合　　計	56kB

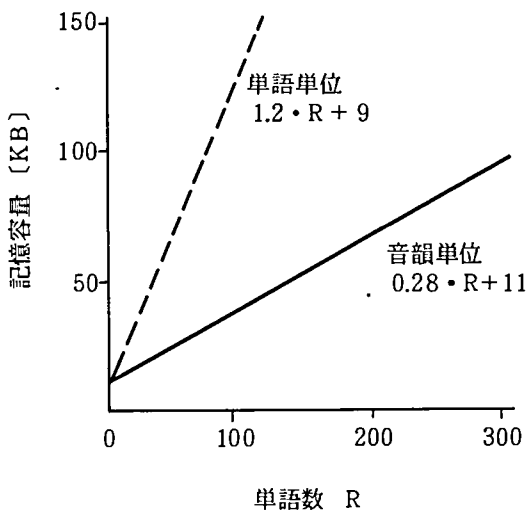


図 11.6　単語数と記憶容量の関係

11.2.4　認識実験

11.2.4.1　認識対象

認識実験により本装置の性能評価を行う。認識対象は表 11.3 に示す，日付（月，日），時刻（時，分），枚数，航空会社名，空港名の 7 種類である。

表 11.3　認識対象

種　類	単語数	内　　容
1. 月	12	1月～12月
2. 日	31	1日～31日
3. 時	15	7時～21時
4. 分	12	0分～55分（5分毎）
5. 枚　数	10	1枚～10枚
6. 航空会社	10	日本航空，日航，全日空，東亜国内航空， 東亜国内，東亜，日本近距離航空， 日本近距離，近距離航空，南西航空，
7. 空港名	67	札幌，旭川，女満別，稚内，釧路，帯広， 函館，秋田，青森，八戸，花巻，東京， 大阪，南紀白浜，新潟，岡山，隠岐， 米子，出雲，徳島，高松，高知，広島， 宇部，松山，大分，福岡，宮崎，鹿児島， 種子島，屋久島，喜界島，奄美大島， 徳之島，沖永良部，山形，仙台，八丈島， 大島，三宅島，富山，金沢，福井， 鳥取，名古屋，北九州，長崎，熊本，福江， 佐渡，壱岐，紋別，中標津，沖縄， 久米島，南大島，宮古，多良間，石垣，与那国 千歳，小松，大村，那覇，三沢，利尻，奥尻

11.2.4.2　音韻標準パターン

認識対象の単語には次の 27 種類の音韻が現われる。

母　　音／a，i，u，e，o／

有声子音／b，d，g，z，dz，r，m，n，$\tilde{g}$，N，j，w／

無声子音／p，t，ts，ch，k，h，f，h_j，s，sh／

一部の子音については調音結合によるスペクトルの変形を考慮に入れ，複数個の音韻標準パターンを用意することとし，実際に用いた音韻標準パターンは表 11.4 に示した 44 種類である。これらについての詳しい説明は第 9 章で述べたのでここでは省略する。

認識実験用の標準パターンは，文献（19）に述べられているものと同様の方法を用いて作成した。すなわち，まず認識対象の単語を 1 回ずつ発声したものを学習サンプルとして，その音韻境界を認識で用いるDPマッチング法によって決める。このときの標準パターンは他の発声者

表 11.4　標準パターンの種類

分　類	数	種　類
母　音	9	a, i, u, e, o, i', u, $\tilde{i}$, $\hat{u}$
母音のわたり	6	ia, io, iu, ai, oi, ui
有声子音	14	b, d, z, dz, $\bar{d}z$, r, $\bar{r}$, m, n, n_j, $\tilde{g}$, N, w, g
無声子音	14	p, t, ts, ch, k_a, k_i, k_u, k_e, k_o, h h_j, f, s, sh
無音区間	1	*
計	44	

の平均的なパターンでよい。次に，この音韻境界をもとにして音韻毎に平均の相関関数を求め，標準パターンを作成する。初期値の標準パターンの影響を軽減するため，この標準パターンを用いて再度学習サンプルの音韻境界を求め直し，最終的な標準パターンを得る。

11.2.4.3　単語辞書

単語辞書として，音韻毎に最小と最大の継続時間の制限をつけた音韻系列を用意する必要がある。音韻の継続時間の制限は学習サンプルの音声から視察により定めた。音韻系列は 1 単語 1 系列が原則であるが，母音の無声化や語頭，語尾の検出もれを考慮し，特定の単語には複数個の音韻系列を用意した。対象別の音韻系列数を表 11.5 に示す。

表 11.5　音韻系列の数

対　象	音韻系列の数
月	36
日	87
時	18
分	12
枚　数	11
航空会社名	10
空港名	110

11.2.4.4 認識結果と検討

男性12名が認識対象の単語を3回ずつ発声した音声を用いて認識実験を行った。発声場所は計算機室である。認識結果を表11.6に示す。航空会社の場合が認識率100％で最も良く，日と

表11.6　誤りの数と認識率

対象＼発声者	NH	IK	SK	SS	FS	KM	NR	KH	WT	HS	SA	KN	計	サンプル数	認識率(%)
1. 月	1	2	0	0	3	3	1	1	2	2	0	3	18	432	93.8
2. 日	3	0	3	3	4	2	0	7	3	2	1	8	36	1116	96.8
3. 時	1	0	6	2	1	1	2	0	3	2	1	3	22	540	95.9
4. 分	0	0	0	1	2	0	0	1	0	1	0	2	7	432	98.4
5. 枚数	0	1	2	0	0	0	1	0	0	0	1	1	6	360	98.3
6. 航空会社名	0	0	0	0	0	0	0	0	0	0	0	0	0	360	100
7. 空港名	3	1	5	6	0	4	1	7	5	3	0	2	37	2412	98.5
計	8	4	16	12	10	10	5	16	13	10	3	19	126		
サンプル数	471	471	471	471	471	471	471	471	471	471	471	471		5652	
認識率(%)	98.3	99.2	96.6	97.5	97.9	97.9	98.9	96.6	97.2	97.9	99.4	96.0			97.8

時の認識率がそれぞれ95.8％，95.9％と低い。平均の認識率は97.8％である。図11.7は候補

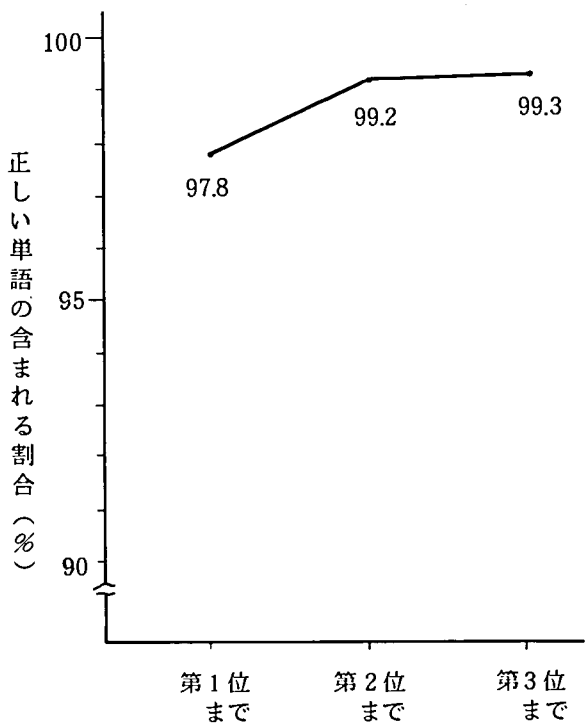


図11.7　候補の数と正しい単語が含まれる割合

の単語を1位から3位までとった場合に，その中に正しい単語が含まれる割合を示したものである。第2位までに99％以上の確率で正しい結果が含まれていることがわかる。また表11.6における空港名の認識結果を第9章における同じ認識対象に対する認識結果と比較してみるとほとんど同じ認識率が得られている。したがって本装置は所期の性能を充分に有しているといえる。

11.3 連続単語音声認識装置 (143)(144)(145)

第10章では，音韻を標準パターンにとった連続単語音声認識法について述べた。そして，3種類の方法を提案し，認識性能，処理量の観点からこれらの方法の比較を行った。一方，連続単語音声認識の応用分野を考えると，計算機へのデータ入力のように認識対象が数字語など語彙数の比較的少ないもので，しかも桁数も3～4である場合が多い。したがって，このような必要条件のもとで，単語音声認識装置と同程度のハードウェア量で連続単語音声認識が可能な装置の開発が望まれる。このような観点から第9章で述べた3種類の方法のうち方法Ⅱ（逆DPマッチング法）を採用し，この方法に基づいた連続単語音声認識装置を開発した。ここでは試作した連続単語音声認識装置の構成と性能について述べる。

11.3.1 連続単語音声認識方法

11.3.1.1 逆DPマッチング法

逆DPマッチング法については第10章で述べたが，装置との関連で再び簡単に説明する。入力音声の特徴パラメータの系列を

$$A = a_1, a_2, \cdots\cdots, a_N \qquad (11.13)$$

とする。式（11.13）の時間軸を逆にして並べた系列を

$$\tilde{A} = a_N, a_{N-1}, \cdots\cdots, a_1 \qquad (11.14)$$

とする。認識対象の単語を1ないし2個つないでできる連続単語に対応する標準パターンの特徴パラメータの系列を

$$B^{\tau}=b^{\tau}_1,\ b^{\tau}_2,\ \cdots\cdots\cdots,\ b^{\tau}_M \qquad (11.15)$$

$$(1 \leqq \tau \leqq R^2+R)$$

とする。ただしRは認識対象の単語数である。式（11.15）の時間軸を逆にした系列を

$$\widetilde{B}^{\tau}=b^{\tau}_M,\ b^{\tau}_{M-1},\ \cdots\cdots\cdots,b^{\tau}_1 \qquad (11.16)$$

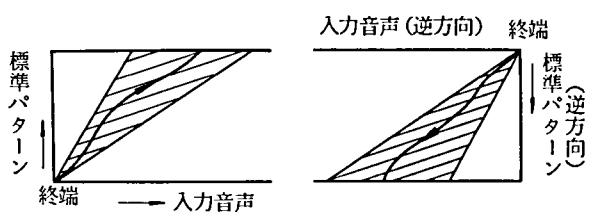
とする。式（11.13）の先頭から第iフレームまでと，式（11.15）の先頭から第jフレームまでのDPマッチングを行い，得られた類似度を

$$L=(i,\ j \mid \tau) \qquad (11.17)$$

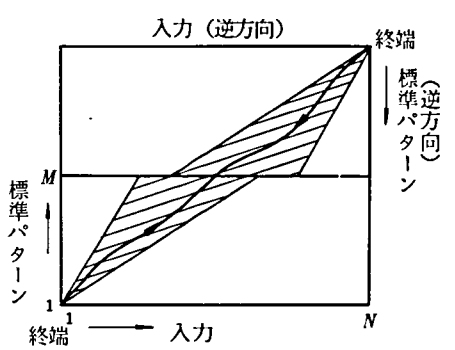
とする。入力の各フレームごとに

$$L(n)=\max_{\tau} L(n,\ M \mid \tau) \qquad (11.18)$$

を満足する類似度$L(n)$，および対応する単語系列$W(n)$を求める。これは図11.8(a)に示すように，式（11.13）と式（11.15）の始端を一致させ式（11.13）の終端を開放した端点フリーDPマッチングを行うことに相当する。



(a) 正方向DPマッチング (b) 逆方向DPマッチング



(c) 正方向と逆方向の結合

図11.8　逆DPマッチングの概念図

同様に，式（11.14）の先頭から第iフレームまでと式(11.16）の先頭から第jルームまでのDPマッチングを行い，得られた類似度を

$$\widetilde{L}\ (i,\ j \mid r) \qquad (11.19)$$

とする。このとき，各フレームごとに

$$\widetilde{L}\ (n) = \max_{r}\ L\ (n,\ M \mid r) \qquad (11.20)$$

を満足する$\widetilde{L}(n)$，および対応する単語列$\widetilde{W}(n)$を求める。これは，図11.8（b）に示すように，入力音声と標準パターンの終端を一致させ，時間軸を逆にさかのぼりながら入力の始端を開放した端点フリーDPマッチングを行うことに相当する。

最後に，式（11.18）と式（11.20）の結果を総合して，

$$\max_{n}\ \left\{\ L\ (n) + \widetilde{L}\ (N-n)\ \right\} \qquad (11.30)$$

を満足するn^{*}を求めると，単語系列$W\ (n^{*})\widetilde{W}\ (N-n^{*})$を認識結果とする。これは，図11.8(c)に示すように，正方向と逆方向のDPマッチングを途中で最適に結合したことに相当する。本方法により，1～4桁の連続単語の認識が可能である。

11.3.1.2 候補単語の制限

処理量を削減するため，処理過程で候補単語を上位から数個にしぼる操作を行う。入力音声（11.13）の先頭から第iフレームまでと標準パターン（11.15）の先頭から第jフレームまでの類似度を$L\ (i,\ j,\ \mid r)$とする。あらかじめ定めた入力の第Iフレームにおいて

$$\max_{m}\ L\ (I,\ m \mid r) \qquad (11.31)$$

を各標準パターンごとに求める。この値が上位数個のものについてのみDPマッチング計算を続け，それ以外については処理を途中で打切る。これは，図11.9に示したように，入力と標準パターンの始端を一致させ，標準パターンの終端を開放した端点フリーDPマッチングを行うことに相当する。

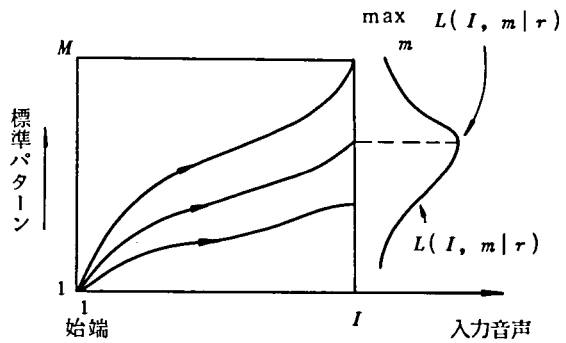


図 11.9　入力音声と標準パターンの
端点フリーDPマッチング

11.3.2　認識装置の構成

本装置は，相関分析に基づいた特徴抽出，音韻標準パターンとの類似度を求める類似度計算，単語系列との類似度を求めるDP計算のように比較的単純で処理量の多い演算が多いため，単語音声認識装置と同様に，主な処理は専用のハードウェアによって行い，処理全体の制御のためにマイクロプロセッサを用いるという構成をとった。装置の構成を図 11.10 に示す。また，

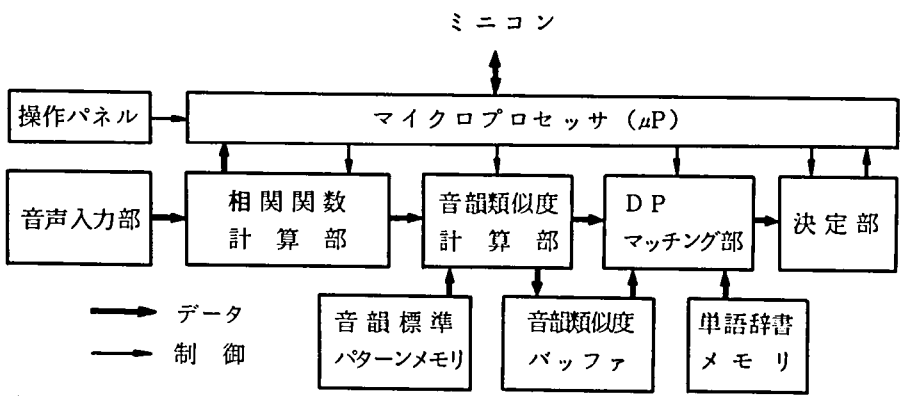


図 11.10　連続単語音声認識装置の構成

外観を図 11.11 に示す。本装置は，入力音声の特徴量として一定長のフレーム毎に自己相関関数を計算する相関関数計算部，標準パターンメモリに蓄えられた音韻標準パターンとの類似度を計算しバッファに蓄える音韻類似度計算部，単語辞書に蓄えられた音韻系列と入力音声の正

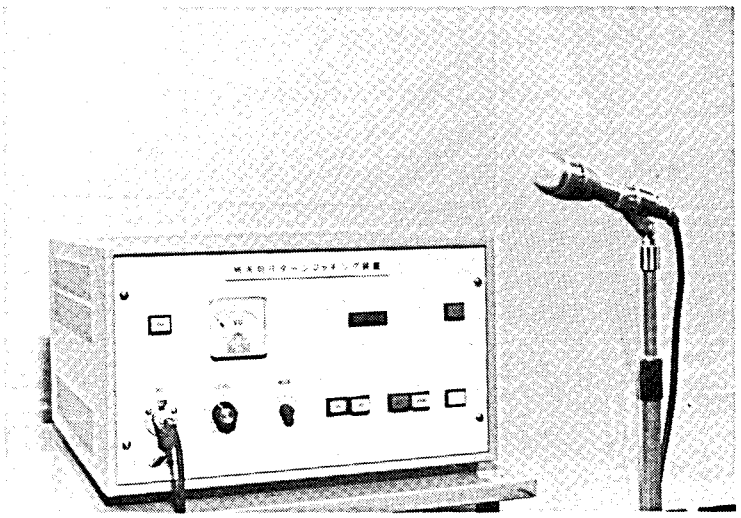


図 11.11　連続単語音声認識装置の外観

方向，および逆方向のDPマッチングを行うDPマッチング部，正方向，逆方向のマッチングの結果を統合して最適の単語系列を認識結果として選択する決定部，音声区間の検出と処理全体の制御を行うマイクロプロセッサ等から構成されている。

図 11.8 からわかるように，本装置は単語音声認識装置とほぼ同じ構成になっている。逆方向DPマッチングのために類似度値を蓄えておく音韻類似度バッファが追加されていることが最大の相違である。

次に各部の動作について述べる。ただし，音声入力部，相関関数計算部については，単語音声認識装置と同一であるため省略する。

11.3.2.1　音韻類似度計算部

15 msec のフレームごとに，各音韻の標準パターンとの類似度を単語音声認識装置と同じ処理により計算する。音韻類似度計算部では最大 63 個の標準パターンに対する類似度が計算できるように設計されている。さらに本装置では単語音声認識装置より細かなパワー情報の利用が可能となっている。すなわち，各音韻に対し 1 ワード，計 63 ワードの ROM が用意されており，各ワードには 1 ビットのフラグと 15 ビットの閾値が書きこまれている。フラグが 1の場合は，音声パワーが閾値以上になると対応する音韻との類似度を低い値に設定する。逆にフラグが 0 の場合は，音声パワーが閾値以下になると，対応する音韻との類似度を低い値に設定する。この機能を利用すると，音声パワーの小さい区間で母音との類似度を小さくしたり，音声パワーの

大きい区間で無声子音との類似度を小さくすることができ，入力と音韻系列の誤った対応づけが行われるのを防ぐことに効果がある。

類似度計算部で得られた入力のフレーム毎の類似度は，DPマッチング部に送られると同時に音韻類似度バッファに蓄えられる。バッファに蓄えられた類似度値は逆方向のDPマッチングを行う際に用いられる。

11.3.2.2 DPマッチング部

単語辞書中には，認識対象の単語の正方向および逆方向の音韻系列が用意されている。2桁の連続単語の音韻系列は，単語の音韻系列を結合することにより自動的に作成される。

正方向のDPマッチングを行うには，音韻類似度計算部から送られてくる音韻類似度系列と，辞書から作成された1～2桁の連続単語の音韻系列との対応づけをDPマッチングにより行う。DPマッチングの具体的な手法は，11.2.2.4で述べた単語音声認識装置における方法と全く同様である。すなわち，音韻系列ごとに式（11.12）に示された個数のワークメモリを用意しておき，その内容を入力音声のフレーム毎に式（11.9a），（11.9b）の漸化式に従って更新してゆけばよい。式（11.8）の右辺の$L(n, M \mid r)$は対応する音韻系列のワークメモリの最終番地に保持されている。したがって，DPマッチング部では，正方向マッチングの際，入力の各フレーム毎に各音韻系列に対応するワークメモリの最終番地に保持されている類似度を取り出してその最大値を選び，その値および音韻系列の種類（これが11.3.1.1で述べた$L(n)$および，$W(n)$に相当する）を決定部へ送ってやればよい。また，式（11.31）の$\max_{r} L(I, m \mid r)$は入力の第Iフレームが入力されたときの，音韻系列rに対応したワークメモリ中の値の最大値である。したがって，候補単語の切りすてを行うには，入力の第Iフレームが入力されたとき，各音韻系列毎にワークメモリ中の最大値を求め，その値が上位数個のものについてのみ，それ以後のDP計算を続ければよい。

次に，逆方向DPマッチングの場合には，音韻類似度バッファから時間軸を逆にして送られてくる類似度系列と，逆方向辞書から作成された1～2桁の連続単語の音韻系列とのDPマッチングを同様の手順で行う。正方向，逆方向のDPマッチングの切り換えは，マイクロプロセッサの制御により行う。

11.3.2.3 決定部

決定部では，DPマッチング部から送られてくる，$L(n)$, $W(n)$, $\tilde{L}(n)$, $\tilde{W}(n)$を蓄えて

おき，正方向，逆方向のDPマッチングが終了した時点で式（11.30）の計算を行い，認識結果の単語系列を求める。

11.3.2.4　マイクロプロセッサによる制御

マイクロプロセッサは，音声の始端，終端の検出，およびDPマッチング部，決定部の制御を行う。相関関数計算部から音声パワーをフレーム毎に受取り，閾値以上かどうかで音声の始端が検出されると，DPマッチング部に正方向マッチングの開始を指示する。音声パワーが閾値以下になると，終端候補と判定して，正方向マッチングの空き時間を利用して逆方向マッチングを開始させる。正方向マッチングは入力と同期して行われるが，逆方向マッチングは入力と非同期に行えるので，正方向マッチングより高速に実行できる。音声パワーが閾値以下の区間が一定時間以上続くと，音声が終了したと判断して，正方向マッチングは終了し，それ以後は逆方向マッチングのみを行う。逆方向マッチングが終了すると，認識結果を求めディスプレイに表示する。図11.12に処理のタイムチャートを示す。音声終了後に逆方向マッチングを行う

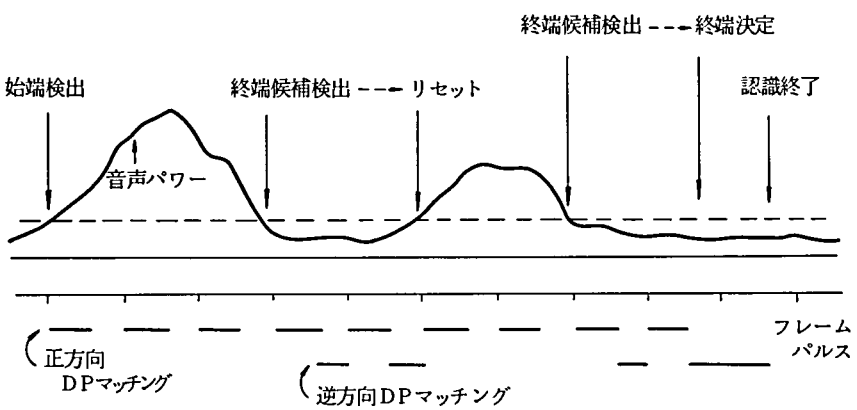


図11.12　連続単語音声認識のタイムチャート

ので，厳密な意味での実時間処理ではないが，上に述べたように逆方向マッチングの一部は正方向マッチングの空き時間を利用して行うため，発声終了後，400～500 msec 程度で結果が得られる。単語音声認識の場合でも，音声が終了したことを確認するのに300～400 msec 程度の遅れが生じることを考えると，これは十分速い認識時間である。

11.3.3 認識装置の仕様

本装置の仕様を表11.7に示す。また，その記憶容量のみを取り出して表11.8に示す。記憶容

表11.7　連続単語音声認識装置の仕様

サイズ	450 W× 250H × 500L　（mm）	
音声入力部	等化器 L. P. F. AD変換器	6 dB／ octの高域強調 f_0 = 3.2 kHz 8 kHz サンプリング 12 ビット量子化
相関関数計算部	フレーム周期 分析次数 相関関数の表現形式	15 ms 0次～10次 v_0 のMSBを1になるようシフト正規化 16 ビット
音韻類似度計算部	標準パターンの表現形式 標準パターン数 類似度の計算方法	最尤スペクトルパラメータを±1の範囲に正規化，16 ビット 最大63個 最尤スペクトル分析法に基づく類似度
DPマッチング部	単語辞書 単語辞書の表現形式 単語辞書の音韻系列 DPマッチング法	正方向，逆方向の2種類 各音韻の最小，最大継続時間を指定した音韻系列 正方向，逆方向共最大32個 逆DPマッチング法 正方向……パイプライン方式による実時間処理 逆方向……正方向マッチングの空き時間を利用
決定部	認識結果	類似度和第1位～第3位の単語系列を選択
マイクロプロセッサ（μp）		M 6800
その他の機能	認識可能桁数 桁数指定	1～4桁 1桁，2桁，3桁，4桁，自由の5つのモードが選択可能

量は，DP計算時のワークメモリが大きいこと，音韻類似度を蓄えるバッファが余分にいることから単語音声認識装置に比較して約2倍になっている。その他のハードウェア量は単語音声

認識装置とほぼ同じであるが，実装方法等によりかなり小型に作ることができた。

表 11.8　連続単語音声認識装置の記憶容量

内　　容		容　　量
標準パターンメモリ		2 kB
音韻類似度バッファ		32 kB
単語辞書メモリ		16 kB
ワークメモリ		48 kB
マイクロプロセッサ用メモリ		4 kB
その他	表示データ	9 kB
	対数値テーブル	
	アドレステーブル	
合　　計		111 kB

本装置の別の特徴として桁数指定機能がある。逆DPマッチング法では，桁数が指定された場合の認識が容易に行える。たとえば，1桁単語の認識は，正方向マッチングとして1桁の標準パターンのみとマッチングを行い，逆方向マッチングは行わない。3桁の連続単語の場合は，正方向は2桁の標準パターンとマッチングし，逆方向は1桁の標準パターンとマッチングすればよい。本装置では，切換スイッチにより，1桁，2桁，3桁，4桁，および桁数自由の5つのモードが選択できるようになっている。

11.3.4　認識実験

本装置の性能評価のために認識実験を行った。連続単語音声認識の応用範囲としては，計算機へのデータ入力等が考えられる。そこで，数字音声を認識対象とした。これは，第9章で述べた計算機シュミレーションによる連続単語音声認識実験の認識対象と同じである。用いた音韻標準パターンを表11.9に示す。また，正方向辞書，逆方向辞書を表11.10，表11.11にそれぞれ示す。表11.9，表11.10はそれぞれ表10.2，表10.3と同じものであり，詳しい説明は省略する。表11.11は表11.10の各音韻系列の時間順序を逆にしたものである。

表 11.9　音韻標準パターンの種類

母　　音	a, i, u, e, o, i′, u′, $\tilde{i}$,
母音のわたり	io, iu,
子　　音	g, r, n_i, N, h, ch, k_i, k_u, s, n,
無音区間	*

表 11. 10　正方向辞書

数　字	音　韻　系　列
1	* i * ch i * i * ch i′ * i * ch
2	* n_i $\tilde{i}$
3	* s a N
4	* i io o N
5	* g o
6	* r o * k_u u * r o * k_u u′ * r o * k_u
7	* n a n a
8	* h a * ch i * h a * ch i′ * h a * ch
9	* k_i i iu u
0	* r e i

表 11.11　逆方向辞書

数　字	音　韻　系　列
1	i ch * i * i′ ch * i * ch * i *
2	$\tilde{i}$ n_i *
3	N a s *
4	N o io i *
5	o g *
6	u k_u * o r * u′ k_u * o r * k_u * o r *
7	a n a n *
8	i ch * a h i′ ch * a h ch * a h
9	u iu i k_i *
0	i e r *

男性発声者が 1 ～ 4 桁の連続数字を各 50 個ずつ発声し認識装置に投入した。全サンプル数は 2200 でこれは 5500 数字に相当する。実験は，騒音レベル 70 dB（A）の計算機室で行った。また認識モードは「桁数自由」に設定した。認識結果を表 11.12 に示す。平均の認識率は 99.3% である。これは第 9 章で述べた計算機シュミレーションによる認識率 $\hat{99.7}$ %に比較して 0.4 % 低下しているが，環境条件の違いを考慮すると妥当な値といえる。したがって本装置は所期の認識性能を持つことが確認された。

表 11.12　装置を用いた認識実験の結果

	Number of errors							Average
	KH	RN	KK	HN	AN	KI	NK	
1 桁 数 字	0	0	0	0	0	1	0	99.7%
2 桁 数 字	0	0	1	2	0	0	1	99.4%
3 桁 数 字	0	0	0	0	2	2	0	99.6%
4 桁 数 字	2	1	2	2	0	4	4	98.9%
Average	99.6%	99.8%	99.4%	99.2%	99.6%	98.6%	99.0%	99.3%

11.4　オンライン会話音声認識システム

11.4.1　システムの構成[94]

第 8 章では会話音声認識システムの構成について述べた。本システムは専用のハードウェアの使用等により実時間の数倍以内で動作するオンラインシステムである。ここでは，そのハードウェア構成，処理の高速化のために取った手法について述べる。

システムの構成を図 11.13 に示す。本システムは，音響処理部，言語処理部，音声応答部から

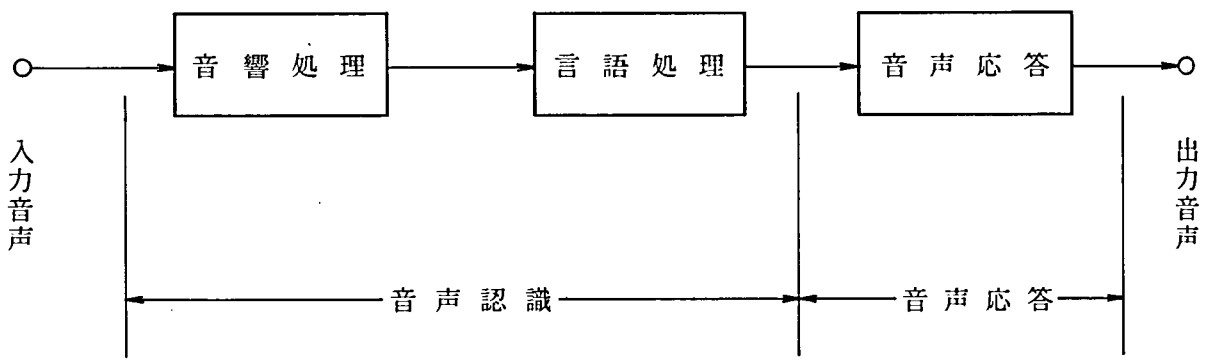


図 11.13　会話音声認識システムの構成

ら構成されている。入力音声の認識は音響処理部，言語処理部で行い，認識結果を音声応答部へ送る。音声応答部では，それに基づいて，問合わせ，確認などの応答文を作成し，それを応答音声で利用者に伝える。このような質問回答のくり返しにより会話音声による各種のサービスが提供できる。

使用する計算機システムの構成を図11.14に示す。音響処理部と音声応答部は同一のミニコ

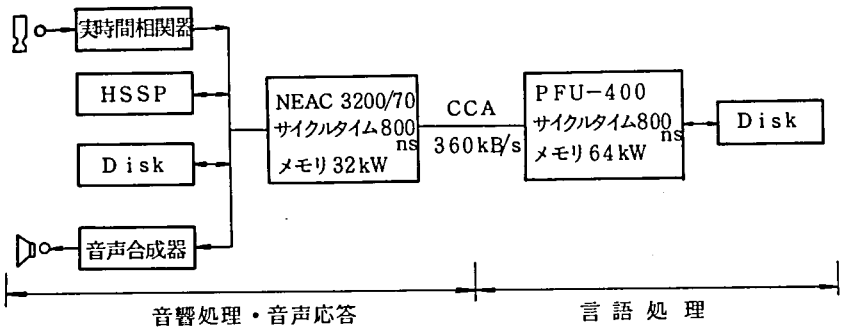


図11.14　会話音声認識システムのハードウェア構成

コン（NEAC 3200 / 70）の上に作成されている。このミニコンは，実時間相関器，音声情報高速処理装置（HSSP），音声合成器が結合されており，利用者とのインターフェースの役割をしている。言語処理部は別のミニコン（PANAFACOM U−400）の上に作成されている。両方のミニコン間では高速のデータ転送が可能であり，音響処理と言語処理の並列処理が行える。音響処理部における処理に特に時間を要するので，以下，音響処理の高速化，全体での処理時間について述べる。

11.4.2　音響処理の高速化[146][147]

音響処理部の構成を図11.15に示す。音響処理部は，特徴抽出と音韻認識の2つの処理が，中心となる。

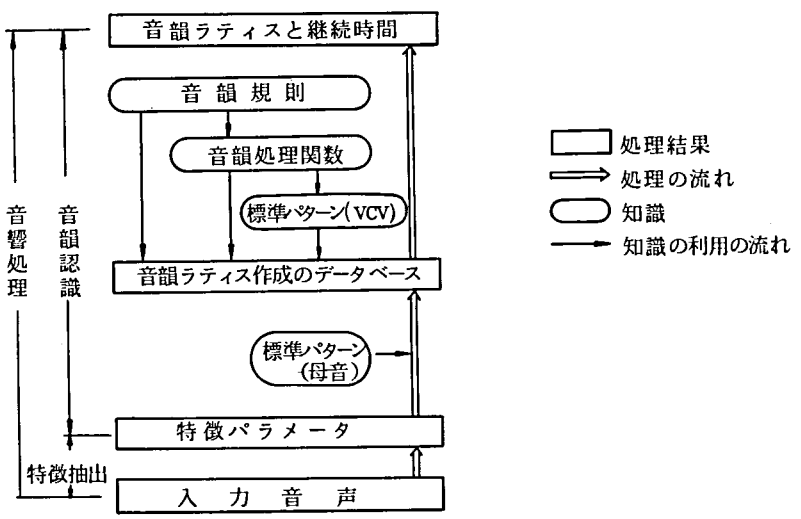


図11.15　音響処理の構成

11.4.2.1 特徴抽出

特徴抽出のフローチャートを図11.16に示す。入力の会話音声は，文節ごとに約1秒のポーズを挿入して発声するものとする。入力音声は，実時間相関器によりリアルタイムで自己相関関数が計算される。次に，母音認識，パワーによるセグメンテーション，音韻ラティスの初期値作成が行われ，結果はディスクに蓄えられる。

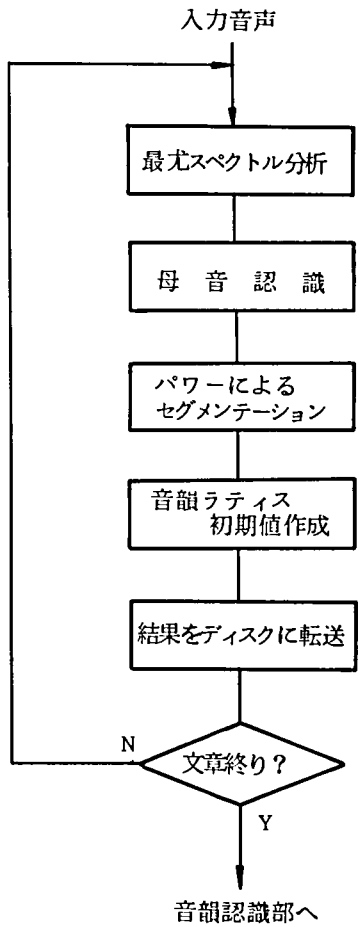


図11.16　特徴抽出のフローチャート

11.4.2.2 音韻認識

音韻認識部では音韻ラティス初期値に各種の処理を施し，最終的な音韻ラティスを作成して言語処理部へ送る。音韻認識で最も時間を要するのはVCV音節単位の子音認識であり，以下に述べる方法でその高速化をはかった。

(1) VCV音節の認識法

VCV音節の認識法について簡単に述べる。入力音声から切り出されたVCV音節は，相関関数の時系列

$$\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots\cdots\cdots, \boldsymbol{v}_N \qquad (11.32)$$

（ただし，$\boldsymbol{v}_i = (v_{i0}, v_{i1}, \cdots\cdots\cdots, v_{ip})$）

とする。VCV音節標準パターンは，最尤スペクトルパラメータの時系列

$$\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots\cdots\cdots, \boldsymbol{A}_M \qquad (11.33)$$

（ただし，$\boldsymbol{A}_j = (A_{j0}, A_{j1} \cdots\cdots\cdots, A_{jp})$）

として蓄えられており，これから類似度マトリクス

$$LM = \{ \iota(i, j,) \} \qquad (11.34)$$

を作成する。ただし

$$\iota(i, j,) = -\log\left\{ \sum_{\tau=0}^{p} A_{j\tau} v_{i\tau} \right\} + \log\left\{ \sum_{\tau=0}^{p} B_{i\tau} v_{i\tau} \right\} \qquad (11.35)$$

を作成する。式（11.35）で（$B_{i0}, B_{i1}, \cdots\cdots B_{ip}$）は$\boldsymbol{v}_i$より得られる最尤スペクトルパラメータである。通常は式（11.35）の第2項を無視した相対的な類似度

$$\iota(i, j) = -\log\left\{ \sum_{\tau=0}^{p} A_{j\tau} v_{i\tau} \right\} \qquad (11.36)$$

を用いる。類似度マトリクスから

$$\max\left\{ \sum_{i=1}^{N} \iota(i, f(i)) \right\} \qquad (11.37)$$

$$\left(\text{ただし，} \iota \leqq f(1) \leqq f(N) \leqq M \quad f(i) = \begin{cases} f(i-1) \\ f(i-1) + 1 \\ f(i-1) + 2 \end{cases} \right)$$

を計算すると，これが両パターン間の類似度である。

(2) 音声情報高速処理装置（HSSP）[151]の利用

HSSPでは高速で浮動小数点演算ができ，かつマイクロプログラム制御が可能であり，類似度計算，DP計算に用いる。図11.17に示すタイムチャートでVCV 音節認識を行う。類似度

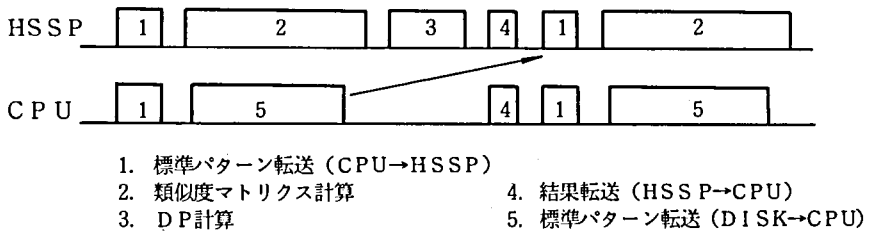


図11.17　VCV音節認識のタイムチャート

計算，DP計算はHSSPで行う。また，標準パターンはディスクに蓄えてあり，処理の高速化のため，類似度マトリクスの計算と，ディスク⟷CPU間の標準パターンの転送を並列に行っている。

(3) 類似度絶対値の利用

拗音の認識の際，相対的な類似度を用いると，候補セグメントの抽出⟶標準パターンとのマッチング⟶拗音の標準パターンとの類似度が最大なら拗音と判定，という手順をふむ。これに対し，式（11.35）で与えられる絶対的な類似度を用いると，候補セグメントの抽出⟶拗音の標準パターンとのマッチング⟶類似度が閾値以上なら拗音と判定，という手順になり，マッチングの回数を減少させることができる。

(4) 継続時間情報の利用

有声子音の多くは継続時間が短い。したがって，有声子音の候補セグメントは，まず継続時間をチェックして閾値以上なら有声子音でないと判定して認識を行わないことにより，マッチングの回数を減少できる。

11.4.3 処理時間[133][147]

音響処理に要する処理時間測定のため各種の長さの文節音声を処理して平均的な処理時間を算出した。第1次システムと上で述べた(2)～(4)の手法を取り入れた第2次システムの処理時間

の比較を表 11.13 に示す。最終的には実時間の 3.4 倍で処理できる。VCV 音節認識に要する時間は 1 音節当り平均 0.4 秒である。

表 11.13　音響処理系の処理時間の比較（入力音声 2 秒）

		第1次システム	第2次システム		
			1	2	3
計算機		FACOM 270 / 20	NEAC 3200 / 70		
HSSPの利用		×	○	○	○
類似度絶対値の利用		×	×	○	○
継続時間の利用		×	×	×	○
処理時間	特徴抽出	40秒	2秒		
	VCV音節認識	90秒	4.76秒	4.31秒	4.07秒
	その他	30秒	0.89秒	0.81秒	0.73秒
	全体	160秒	7.65秒	7.12秒	6.80秒
	処理時間／入力音声長	80倍	3.83倍	3.56倍	3.40倍

言語処理部の処理時間は実時間の 2.0 倍であり，音響処理，言語処理をあわせると実時間の 5.0 倍となる。入力音声を文節毎に音響処理と言語処理を並列に行いながら，認識して行く例を図 11.18 に示す。なお，音声応答部では，1 つの応答文を作るのに 2 ～ 10 秒かかる。

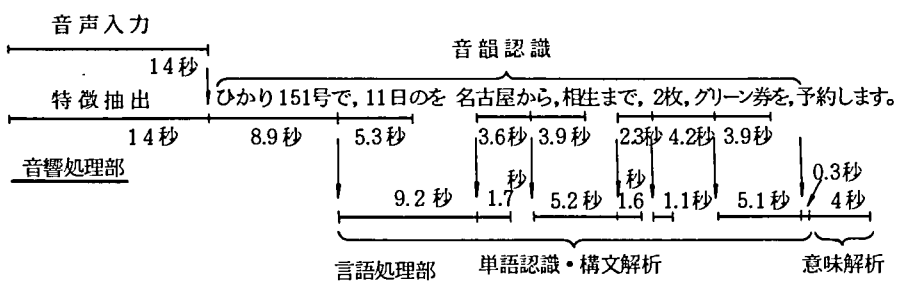


図 11.18　会話音声認識系のタイムチャート

11.5　あとがき

本節では音声認識システムの高速化，ハードウェア化について述べた。まず，実時間で動作する単語音声認識装置について述べた。本装置は，第 9 章で述べた音韻を単位とした単語音声認識法に基づいた装置である。本装置では，相関関数計算，類似度計算，DP 計算のように比

較的単純で処理量の多い演算が多いことから，主な処理は専用のハードウェアで行い，処理全体の制御をマイクロプロセッサで行うという構成をとった。さらに，DPマッチングの方法として実時間処理にむいた方法を取り入れた。これらの結果，実時間で100語以上の単語が認識可能な単語音声認識装置を試作することができた。また認識実験により，所期の性能を持つことを確めた。

次に，連続単語音声認識装置について述べた。本装置では，単語音声認識装置と同じように，主な演算は専用のハードウェアで行い，マイクロプロセッサで全体の制御を行うという構成をとった。連続単語音声の認識法としては，第10章で提案した3種類の方法のうち，装置の応用範囲を考慮して，逆DPマッチング法を採用した。その結果，単語音声認識法とほぼ同じ程度のハードウェア量で，4桁以内の連続単語を実時間に近い処理時間で認識できる装置を試作することができた。また，連続数字音声を用いた認識実験により，本装置が所期の性能を持つことを確めた。

最後にオンライン会話音声認識システムについて述べた。本システムは，第8章で述べた会話音声認識システムの高速化，オンライン化を計ったものである。音響処理部と言語処理部を別のミニコン上に作成し，両者間で高速のデータ転送を行えるようにして，音響処理部と言語処理部の並列動作を可能にした。さらに音響処理については，音声情報高速処理装置，実時間相関器等の専用ハードウェアを使用することにより特徴抽出，音韻認識の高速化を計った。その結果，音響処理部では実時間の3.4倍，言語処理部では実時間の2.0倍，全体では実時間の5.0倍で動作するオンライン会話音声認識システムを作成した。

第12章　結　　言

本論文では，著者が日本電信電話公社武蔵野電気通信研究所および横須賀電気通信研究所で行ってきた日本語音声の自動認識システムについて述べた。

本論文は大きく分けて2つの部分から構成されている。前半では，母音一子音一母音よりなるVCV音節を認識の単位とする音声認識法について述べた。VCV音節を認識の基本単位として，採用したのは次の理由による。

(1)　VCV音節を単位としてセグメンテーションを行うには，母音部分を検出すればよい。音声中の母音部分は比較的容易に検出できる。したがってVCV音節を単位にとることにより，セグメンテーションが行いやすい利点がある。

(2)　連続音声中の子音の認識は困難な問題である。これは，子音が前後の母音の影響を受けて変形しやすいことによる，VCV音節は，子音をある時間点におけるスペクトル情報として表現するのではなく，前後の母音と一緒にしてスペクトルの遷移として動的に表現しているため，子音の認識に有利である。

(3)　日本語は，VCVC……という音韻構成になっているから，VCV音節を音声を構成している基本的な単位であると考えても矛盾はない。

そして対象を単語音声，会話音声と拡大して行きながらＶＣＶ音節を単位とした音声認識法の有効性を示した。

本論文の後半では，音韻を単位とした音声認識法について述べた。これは，VCV 音節を単位にした認識法に対してより実用的な方法の提供を目的としたものである。音韻を単位にした音声認識法の利点は次の通りである。

(1)　本方法は，音韻単位の標準パターンと，音韻系列として表現された単語辞書を併用して認識を行う方法であり，誤りの生じやすいセグメンテーションの操作を必要としないため，高い認識率が得られる。

(2)　認識対象の語彙の増加，変更の際は単語辞書の内容のみを変更すればよいため，認識対象語の増加，変更が容易である。

(3)　通常，発声者が変る際には，認識対象語を全部発声することにより標準パターンを入れかえる操作を行う。この操作は，認識対象語が多くなるとやっかいである。これに対し，音韻を標準パターンの単位にとる方法では，一部の単語を発声しただけで標準パターンを作り変えることが可能であるため，標準パターンの登録の手間が簡単化できる。

(4) 標準パターン，単語辞書を蓄えるための記憶容量が少なくてすむ。

本方法は，単語音声，連続単語音声に適用しその有効性を示した。

以下，各章の内容と得られた成果について簡単にまとめておく。

第1章では，音声認識全般にわたる研究動向について述べるとともに，音声認識研究の問題点をあげ，研究の基本方針および本論文の構成について述べた。

第2章では，本論文で採用した音声分析法である最尤スペクトル分析法について述べると共に，パターンマッチングの際用いる距離尺度（類似度尺度）を明らかにした。

第3章から第6章まではVCV音節を単位とした音声認識法について述べた。

第3章では，VCV音節の認識法について述べた。内容および得られた成果をまとめると次のようになる。

(1) 特徴パラメータの時系列で表現されるVCV音節の学習サンプルから，発声のばらつきを吸収した平均化された標準パターンを作成する方法を提案し，その有効性を示した。

(2) 音韻，VC・CV音節等を単位にした標準パターンと比較して，VCV音節を単位とした標準パターンの方が子音の認識において優れている事を示した。

(3) パワー情報を補助情報として使うことにより，88.0％の子音認識が得られた。

第4章では，VCV音節を単位とした単語音声認識について述べ，次のような成果が得られた。

(1) 入力音声をVCV音節単位にセグメンテーションし，次に各セグメントの認識を行い最後に，VCV音節の系列として記述してある単語辞書との照合により単語音声を認識する手法を提案し，有効性を示した。

(2) セグメンテーションの方針として，セグメンテーション時の誤りが致命的となりやすい事を考慮し，あいまいさを許したセグメンテーションを行うこととした。この方針は，連続単語音声認識，会話音声認識にも引き継がれた。

第5章では，対象を連続単語音声に拡張し，VCV音節を単位とした連続単語音声認識について述べた。連続単語音声は，発声者の負担を軽減し，情報発生速度を速くできるという意味で実用的見地から有利であると同時に，単語音声認識から会話音声認識へ進む際の中間的なステップとしても適当である。連続単語音声認識では，単語単位のセグメンテーションをどうするかという問題が生じる。それに対しここでは，すべての単語系列と入力音声との類似度をDPを用いて求め，最も大きい類似度を持つ単語系列を認識結果とする方法を提案した。この方法は，セグメンテーションを必要としないので高い認識率が得られ，かつ，DPを用いることにより比較的少ない処理量ですむという利点がある。この方法は，現在では，連続単語音声認識の基

本手法として定着している。

第6章ではVCV音節の認識法の改良について述べた。第4章，第5章で述べた単語音声認識連続単語音声認識の性能をさらに向上させるには，VCV音節の認識性能の向上が必要である。ここでは，問題を(1)単独のVCV音節のように，母音定常部を含んだVCV音節の認識，(2)連続音声中のVCV音節のように，母音定常部を含まないVCV音節の認識，の2つに分け，それぞれについて論じた。その結果，次のような事実が明らかになった。

(1) 母音定常部を含んだVCV音節の認識の際は，子音部分，および母音から子音への遷移部分に重みをつけたパターンマッチング法が有効である。

(2) 母音定常部を含まないVCV音節の認識には，入力を標準パターンの一部に対応づける端点フリーDPマッチング法が有効である。

さらに端点フリーマッチング法を連続単語音声認識に適用すると，認識率の向上が見られ，有効性が確められた。

第7章では会話音声を対象とした音声認識システムについて音響処理を中心に述べた。本システムの特徴は次の点である。

(1) 一語一語の認識を目ざすのではなく，入力音声から意味内容を抽出する音声理解システムの構築を目指した。

(2) システムと発声者が質問回答形式で対話を進めながら，発声者の意図をシステムが理解する，いわゆる質問回答システムの構成をとった。

(3) 音響処理結果の表現形式として，セグメンテーション，音韻認識のあいまいさを含めた音韻ラティスを採用した。

(4) 音響処理の構成方針としては，特徴パラメータに bottom－up 的な処理を施して，音韻ラティスを作成することにした。

新幹線の座席予約サービスをタスクとしてシステムを作成し評価を行った。その結果，次のような点が明らかになった。

(1) 音韻ラティスは音韻系列に比較して有利な表現形式である。

(2) システム全体の性能を向上させるには，鼻音，拗音，半母音，連続母音，無声化母音等の処理を正確に行う必要がある。

第8章では，第7章の結果に基づき，第1次システムに改良を加えて作成した会話音声認識の第2次システムについて述べた。本システムは，第1次システムに比較して以下の点に考慮を払っている。

(1) 音声学的な知識を規則化して音韻規則として蓄えておき，これをtop－down 的に利用して音響処理を行う方針をとった。

(2) 比較的学習の容易な母音標準パターンについては発声者毎に用意し，VCV 音節標準パターンは固定しておくことにより，新しい発声者に対応し易いシステム構成をとった。

(3) 会話音声認識システム全体の高速化，オンライン化をはかり，実際の質問回答実験によりシステムの評価が行えるようにした。

本システムの性能評価実験により，次のような点が明らかになった。

(1) top-down 的な処理を取り入れることにより，音響処理部の構成を柔軟性に富んだものにすることができた。

(2) 言語処理を含めたシステム全体の認識実験により，今後重点的に改良を加えてゆくべき点が明らかになった。

(3) VCV音節標準パターンは固定し，母音標準パターンのみ学習するという方式で,十分発声者の交替に追従できることが明らかになった。

(4) 質問回答形式を取ることにより，極めて高いタスク完了率を得ることができる。また，認識率の低い発声者に対しても，サービス性はあまり低下しないことが明らかになった。

第9章から第10章までは音韻を単位とした音声認識法について述べた。

まず，第9章では音韻を単位とした単語音声認識法について述べた。ここでは，音韻単位の標準パターンと音韻系列として表現されている単語辞書との併用により単語音声認識を行った。単語辞書の表現方法として次の2つを提案した。

(1) 辞書Ⅰは，音韻系列の各音韻に平均的な継続時間情報が付加されたものである。

(2) 辞書Ⅱは，各音韻に最小，最大の継続時間情報が付加されたものである。

単語辞書の表現形式に応じて，DPマッチングについても方法Ⅰ，Ⅱの2つの方法を提案した。さらに，標準パターンについては，日本語に現われる音韻をベースにした標準パターンⅠと，それに母音間のわたり，子音の調音結合を考慮した擬似的な音韻を追加した標準パターンⅡの2種類を用意した。航空機の座席予約サービスを想定した169単語を対象とした認識実験により次の点が明らかになった。

(1) 認識性能は方法Ⅱの方が優れている。方法Ⅱは実時間処理に適した方法であるため，認識装置を作成する際には方法Ⅱが適している。

(2) 標準パターンについては，標準パターンⅡの方が高い認識率を示した。したがって，辞書の作成に際しては実際の音声現象を考慮することが必要である。

第10章では音韻を単位とした連続単語音声認識について述べた。以下の特徴を持つ3種類の方法を提案した。

(1) 方法Ⅰは第5章で述べた連続単語音声認識法と音韻単位の音声認識法を結合，発展させたものである。入力音声とすべての単語系列のマッチングをDPを用いて行い，最大の類似度を持つ単語系列を認識結果とする方法である。

(2) 方法Ⅱは，通常のDPマッチングに加え，音声の終端から時間軸を逆方向にしたDPマッチングを行い，これらを合わせて入力音声と単語系列の最適なマッチングを行う方法である。

(3) 方法Ⅲは入力音声と単語とのマッチングを，第6章で提案した端点フリーDPマッチングで行って，入力音声と単語との類似度を求めた後，音声の終端から1語ずつ単語を求めて答を得る方法である。

数字音声を対象とした認識実験，および処理量の比較により各方法の評価を行った。その結果，次の事実が明らかになった。

(1) 方法Ⅰは認識性能が最も高い。

(2) 方法Ⅱは認識性能はⅠに匹敵し，処理量が少ない。ただし，単語数，桁数の制限を受ける。

(3) 方法Ⅲは認識性能は少し劣るが，処理量は最も少ない。

第11章では，以上述べてきた日本語音声認識システムの終大成として，システムの高速化，ハードウェア化について述べた。まず単語音声認識装置について述べた。本装置は第9章で提案した方法Ⅱに基づいて試作されたものであり，200語程度を対象とした実時間処理が可能である。次に，連続単語音声認識装置について述べた。本装置は第10章で提案した方法Ⅱに基づいて試作したもので30語を対象とした4桁以内の連続単語音声の認識が可能である。これらの装置はいずれも，性能評価実験により所期の性能を満足することが確認された。最後にオンライン会話音声認識システムについて述べた。これは，第8章で述べた会話音声認識システムの高速化，オンライン化を計ったもので，音声処理における音声処理用ハードウェアの使用，音響処理と言語処理の並列処理等により，実時間の5倍以内の処理速度を達成した。

以上述べてきたように，本論文は単語音声から会話音声に至る極めて広い範囲の音声認識法について論じたものである。しかも，基礎的な方法のみならず実用をめざした方法についても述べており，認識装置についても言及している。本論文の研究成果は，音声認識の基礎的な技術レベルの向上，および，実用化の促進に寄与する所が極めて大きいものと確信する。

謝　　辞

本研究の開始以来，長期間にわたって当所基礎研究部第四研究室長及び斎藤特別研究室長として御指導を賜った東京大学教授・斎藤収三博士に心から感謝の意を表します。また，研究の経過や論文のとりまとめにおいて種々の面から御教示を賜った京都大学教授・坂井利之博士に深く感謝いたします。

また，この研究の機会を与えられ，また御鞭達下さった横須賀電気通信研究所宅内機器研究部長・松田亮一博士，武蔵野電気通信研究所基礎研究部長・畔柳功芳博士，宅内機器研究部統括役・苗村明氏，基礎研究部統括役・橋本新一郎博士，宅内機器研究部音声入出力方式研究室長・寺井正明博士，基礎研究部第4研究室長・板倉文忠博士に感謝の意を表します。また，本研究は直接指導を頂いた武蔵野電気通信研究所研究専門調査役・好田正紀博士の適切な御指導および激励なくしては進展しえなかったものであり深く感謝いたします。

さらに，日頃有益な討論をして下さった横須賀電気通信研究所研究専門調査役石井直樹氏，研究専門調査員・長島広海氏，武蔵野電気通信研究所研究専門調査役・古井貞熙博士，研究専門調査員・鹿野清宏博士ならびに宅内機器研究部音声入出力方式研究室，基礎研究部第四研究室の諸氏に厚くお礼申し上げます。

文　　献

(1) 新美康永：＂音声認識＂，情報科学構座，E－19－3，共立出版（1979）

(2) 中田和男：＂音声＂，日本音響学会編，音響工学講座7，コロナ社（1977）

(3) 斎藤，中田：＂音声情報処理の基礎＂，オーム社（1981）

(4) 電子通信学会編：＂聴覚と音声＂，電子通信学会，コロナ社（1980）

(5) J.L. Flanagan: "Speech Analysis, Synthesis and Perception", 2nd Edition, Springer-Verlag (1972)

(6) J.L. Flanagan: "Computers That Talk and Listen; Man-Machine Communication by Voice", Proceeding of the IEEE, Vol. 64, No.4 (1976-04)

(7) D.R. Reddy: "Speech Recognition by Machine; A Review", Proceeding of the IEEE, Vol. 64, No.4 (1976-04)

(8) A. Newell, J. Barnet, J.W. Forgie, C. Green, D. Klatt, J.C.R. Licklider, J. Munson, D.R. Reddy and W.A. Woods: "Speech Understanding System - Final Report of a Study Group", North-Holland (1973)

(9) D.H. Klatt: "Review of the ARPA Speech Understanding Project", Journal of the Acoustic Society of America, Vol. 62, No.4 (1977-06)

(10) W.A. Lea and J.E. Shoup: "Gaps in the Technology of Speech Understanding", Draft of Oral Presentation at IEEE International Conference on ASSP(1978)

(11) K.H. Davis, R. Biddulph and S. Balashek: "Automatic Recognition of Spoken Digits", Journal of the Acoustic Society of America, Vol. 24, p.637 (1952-10)

(12) H. Dudley and S. Balashek: "Automatic Recognition of Phonetic Patterns in Speech", Journal of the Acoustic Society of America, Vol. 30, p.721 (1958-08)

(13) E.E. David: "Artificial Auditory Recognition in Telephony", IBM Journal, Vol.2, p.294 (1958-08)

(14) P. Denes and M.V. Mathews: "Spoken Digit Recognition Using Time-Frequency Pattern Matching", Journal of the Acoustic Society of America, Vol. 32, p.1450 (1960-11)

(15) R.F. Purton: "Speech Recognition Using Autocorrelation Analysis", IEEE Trans. on AU, AU-16, p.235 (1968-06)

(16) 板倉，斎藤：＂統計的手法による音声スペクトル密度とホルマント周波数の推定＂，信学論(A)，53－A，p.35（1970－01）

(17) 板倉文忠：“統計的手法による音声の特徴抽出”，東北大通研音声情報処理シンポジウム，Ⅱ－5（1971－02）

(18) B.S. Atal and S.L. Hanauer: "Speech Analysis by Linear Prediction of the Speech Wave", Journal of the Acoustic Society of America, Vol. 50, p.637 (1971-02)

(19) 好田, 橋本, 斎藤：" 数字音声の機械認識系 ", 信学論(D), 55-D, p.186 (1972-03)

(20) J.H. Warren: "A Pattern Classification Technique for Speech Recognition", IEEE Trans. on AU, AU-19, p.281 (1971-12)

(21) A. Ichikawa, Y. Nakano and K. Nakata: "Evaluation of Various Parameter Sets in Spoken Digits Recognition", IEEE Trans. on AU, AU-21, p.202 (1973-06)

(22) E. Itakura: "Minimum Prediction Residual Principle Applied to Speech Recognition", IEEE Trans. on ASSP, ASSP-23, p.67 (1975-02)

(23) 中津, 好田：" ＶＣＶ音節を単位とした単語音声の認識 ", 信学論(A), 61-A (1978-05)

(24) 中津, 好田：" 連続して発声した単語音声の認識 ", 信学論(D), 61-D(1978-09)

(25) J.H. King and C.J. Tunis: "Some Experiments in Spoken Word Recognition", IBM Journal, Vol. 10, p.65 (1966-01)

(26) G.L. Clapper: "Automatic Word Recognition", IEEE Spectrum, Vol.8, p.57 (1971-08)

(27) P.N. Sholtz and R. Bakis: "Spoken Digit Recognition Using Vowel-Consonant Recognition", Journal of the Acoustic Society of America, vol.34, p.1 (1962-01)

(28) 鈴木, 中田：" 数字語識別の実験 ", 信学誌, Vol.45, p.303 (1962-03)

(29) L. Gill and A.R. Meo: "Sequential System for Recognizing Spoken Digits in Real Time", ACOUSTICA, Vol.19, p.38 (1967/1968)

(30) 加藤, 千葉, 永田：" 数字音声認識装置 ", 信学誌, Vol.47, p.1319 (1964-09)

(31) B. Gold: "Word-Recognition Computer Program", MIT Tech. Rep., 452, p.1 (1966-06)

(32) T.G. von Keller: "An On-Line Recognition System for Spoken Digits", Journal of the Acoustic Society of America, Vol. 49, p.1288 (1971-04)

(33) V.M. Velichko and N.G. Zagoruyko: "Automatic Recognition of 200 Words", Int. J. Man-Machine Studies, Vol. 2, p.223 (1970).

(34) 迫江, 千葉：" 動的計画法を利用した音声の時間正規化に基づく連続単語認識 ", 音学誌, Vol.27, p.483 (1971-09)

(35) 好田, 橋本, 斎藤：" 時間長の伸縮に適応できる各種音声認識法について ", 音学講論, 2-1-14 (1971-05)

(36) G.M. White and R.B. Neely: "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering and Dynamic Programming", IEEE Trans. on ASSP, ASSP-24, p.183 (1976-04)

(37) H.Sakoe and S. Chiba: "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. on ASSP, ASSP-26 (1978-02)

(38) 古井貞熙：“単語音声認識における時間正規化方法の検討”，信学技報，PRL76−25 (1976−07)

(39) P. Denes: "The Design and Operation for the Mechanical Speech Recognition at University College London", J. Brit. I.R.E., Vol.19, p.219 (1959-04)

(40) 樽松，武田，井上：“書換え規則を用いて音声の認識を行なう場合の規則の一講成法” 信学論(D)，55−D，p.91 (1972−02)

(41) 板橋，城戸：“辞書と音形規則を利用した単語音声の認識”，音学誌，Vol.27, p.473 (1971−09)

(42) R.Alter: "Utilization of Contextual Constraints in Automatic Speech Recognition", IEEE Trans. on AU, AU-16, p.6 (1968-03)

(43) J.P. Haton: "A Practical Application of a Real-Time Isolated-Word Recognition System Using Syntactic Constraints", IEEE Trans. on ASSP, ASSP-22, p.416 (1974-12)

(44) S. Rivoira and P. Torasso: "Syntax-Directed Recognition of Spoken Words in Real-Time", 1977 IEEE International Conference on ASSP, p.475 (1977)

(45) S.E. Levinson, A.E. Losenberg and J.L. Flanagan: "Evaluation of a Word Recognition System Using Syntax Analysis", 1977 IEEE International Conference on ASSP, p.483 (1977)

(46) 好田，鹿野：“構文情報を利用した算術式の音声認識”，信学技報，EA73−54 (1974−03)

(47) S. Chiba: "Spoken Word Recognition by Multiple Linear Separation", 6th ICA, B-4-4 (1968)

(48) M.R. Sumbur and L.R. Rabiner: "A Speaker-Independent Digit-Recognition System", BSTJ, Vol.54, p.81 (1975-01)

(49) M.B. Herscher and R.B. Cox: "An Adaptive Isolated-Word Speech Recognition System", 1972 Conference on Speech Communication and Processing, C1 (1972)

(50) 好田，橋本，斎藤：“不完備型の学習サンプルによる音声の認識について”，信学技報，IT71−61 (1971−10)

(51) 坂井，中川，林：“音韻スペクトルの予備学習による限定語彙単語音声の認識”，信学技報，EA75−61 (1976−01)

(52) 古井貞熙：“単語音声認識における学習の能率化”，音声研究会資料，S77−43 (1977−12)

(53) 長島，中津：＂単語音声認識における音韻標準パターンの適応化の検討＂，音声研究会資料，S79－51（1979－12）

(54) 中川，山尾，神谷，坂井：＂音声パターン変動の個人差の正規化法＂，音声研究会資料，S79－53（1979－12）

(55) 松本，脇田：＂Frequency Warping による話者正規化＂，音学講論，3－2－6（1979－06）

(56) L.R. Rabiner, S.E. Levinson, A.E. Rosenberg and J.G. Wilpon: "Speaker Independent Recognition of Isolated Words Using Clustering Techniques", IEEE Trans. on ASSP, ASSP-27, p.336 (1979)

(57) S.E. Levinson, L.R. Rabiner, A.E. Rosenberg and J.G. Wilpon: "Interactive Clustering Techniques for Selecting Speaker Independent Reference Templates for Isolated Word Recognition", IEEE Trans. on ASSP, ASSP-27, p.134 (1979)

(58) P.B. Scott: "VICI-A Speaker Independent Word Recognition System", 1976 IEEE International Conference on ASSP, p.210 (1976)

(59) 坂井，中川：＂不特定話者・連続音声向き単語音声の識別＂，情報処理，Vol.17，p.650（1976－07）

(60) 千葉，亘理，渡辺：＂不特定話者を対象とした単語音声認識システム＂，昭52信学情報全大，219（1977－08）

(61) 長島，中津，小島，石井：＂不特定話者電話音声認識法＂，信学会情報・システム全大，107（1981－10）

(62) 畑中，他：＂細分類音種を用いた不特定話者数字音声の認識＂，音学講論，1－1－21（1980－10）

(63) 山田，能勢　他：＂不特定話者単語音声認識＂，音学講論．2－1－10（1981－10）

(64) J.J. Kalinowski, J.C. Brown and S.G. Bhanji: "Application of Discrete Word Recognition and Response to Multiuser Tactical Communications", 1976 International Conference on ASSP, p.222 (1976)

(65) 鹿野清宏：＂大語い単語音声認識におけるＬＰＣスペクトル　スッチング尺度の評価＂，音学講論，1－1－10（1980－10）

(66) J.L. Flanagan: "Computers that Talk and Listen: Man-Machine Communication by Voice", Proceedings of the IEEE, Vol.64, p.405 (1976-04)

(67) 三輪，牧野，松岡，城戸：＂オンライン単語音声自動認識装置＂，信学技報，PRL76－8（1976－04）

(68) 長島．中津：＂単語音声認識を用いた質問回答システム＂，音声研究会資料，S78－52（1978－12）

(69) L.R. Rabiner and M.R. Sambur: "Some Preliminary Experiment in the Recognition of Connected Digits", IEEE Trans. on ASSP, ASSP-24, p.170 (1976)

(70) 中津，好田：＂ＶＣＶ音節を単位とした連続単語音声の認識＂，信学技報，PRL75−44（1975−10）

(71) H. Sakoe: "Two-Level DP-Matching - A Dynamic Programming Based Pattern-Matching Algorithm for Connected Word Recognition", IEEE Trans. on ASSP, ASSP-27, p.588 (1979-12)

(72) 中津，浜田，石井，高浜：＂日本語単音節音声認識法の検討＂，信学会情報・システム全大，117（1981−10）

(73) 新田，村田，坪井，近藤：＂単音節音声認識の一方法＂信学技報，EA80−62(1980)

(74) 吉田，迫江，千葉：＂日本語単音節音声認識実験＂，音学講論，3−2−16（1979−06）

(75) 坂井，堂下：＂会話音声識別装置＂，信学誌，Vol.46，p.1696（1963−11）

(76) H.F. Olson and H. Belar: "Phonetic Typewriter", Journal of the Acoustic Society of America, Vol. 28, p.1072 (1956-11)

(77) D.R. Reddy: "Computer Recognition of Connected Speech", Journal of the Acoustic Society of America, Vol. 42, p.329 (1967-08)

(78) C.C. Tappert, W.D. Chapman, N.R. Dixon and A.B. Rabinowitz: "Application of Sequential Decoding for Converting Phonetic to Graphic Representation in Automatic Recognition of Continuous Speech (ARCS)", IEEE Trans. on AU, AU-21 (1973-06)

(79) R. Alter: "Utilization of Contextual Constraints in Automatic Speech Recognition", IEEE Trans. on AU, AU-16 (1968-03)

(80) W.A. Woods: "Transition Network Grammars for Natural Language Analysis", Commun. ACM, Vol. 13 (1970-10)

(81) T. Winograd: "Understanding Natural Language", Academic Press (1972)

(82) D.R. Reddy, L.D. Erman and R.B. Neely: "A Model and a System for Machine Recognition of Speech", IEEE Trans. on AU, AU-21, p.229 (1973-06)

(83) V.R. Lesser, R.D. Fennell, L.D. Erman and D.R. Reddy: "Organization of the HEARSAY II Speech Understanding System", IEEE Trans. on ASSP, ASSP-23, p.11 (1975-02)

(84) J.K. Baker: "The DRAGON System - An Overview", IEEE Trans. on ASSP, ASSP-23, p.24 (1975-02)

(85) B.T. Lowerre: "The HARPY Recognition System", Ph.D. Dissertation, Carnegie-Mellon University (1976)

(86) W.A. Woods: "Motivation and Overview of SPEECHLIS: An Experimental Prototype for Speech Understanding Research", IEEE Trans. on ASSP, ASSP-23, p.2 (1975-02)

(87) J.J. Wolf and W.A. Woods: "The HWIM Speech Understanding System", 1977 IEEE International Conference on ASSP, p.784 (1977)

(88) H.B. Ritea: "Automatic Speech Understanding System", COMPCON-75, Fall, p.319 (1975)

(89) D.E. Walker: "The SRI Speech Understanding System", IEEE Trans. on ASSP, ASSP-23, p.397 (1975-02)

(90) W.A. Lea, M.F. Medress and T.E. Skinner: "A Prosodical Guided Speech Understanding Strategy", IEEE Trans. on ASSP, ASSP-23, p.30 (1975-02)

(91) J.W.Klovstad, L.F. Mondshein: "The CASPERS Linguistic Analysis System", IEEE Trans. on ASSP, ASSP-23, p.118 (1975-02)

(92) 中津, 好田："会話音声の機械認識における音響処理", 信学論(D), 61−D, 4, (1978−04)

(93) 鹿野, 好田："会話音声の機械認識における言語処理", 信学論(D), 61−D, 4, (1978−04)

(94) 好田, 中津, 鹿野, 伊藤："音声によるオンライン質問回答システム", 音学誌, Vol. 34, p.194 (1978−03)

(95) S. Nakagawa: "A Machine Understadning System for Spoken Japanese Sentences",
京都大学博士論文 (1976−10)

(96) 新美, 小林, 浅見, 三木："「SPOKEN BASIC1」の認識システム", 情報処理, Vol.5 (1977−05)

(97) 関口, 重永："日本語文章の音声認識システム", 音学誌, Vol.34 (1978−03)

(98) F. Jelinek: "Continuous Speech Recognition by Statistical Methods", Proceedings of IEEE, Vol.64 (1976-04)

(99) G. Mercier, P. Quinton and R. Vives: "Man-Machine Dialogue with KEAL", Centre National d'etudes des Telecommunications, Vol.4 (1977)

(100) R. De Mori, S. Rivoira and A. Serra: "A Speech Understanding System with Learning Capability", Proc. 4th Int. Joint Conf. on AI , p.468 (1975)

(101) J.P. Haton and J.M. Pierrel: "Organization and Operation of a Connected Speech Understanding System at Lexical, Syntactic and Semantic Levels", 1976 IEEE International Conference on ASSP, p.430 (1976)

(102) L.R. Bahl, R. Bakis, P. Cohen, A.G. Cole, F. Jelinek, B.L. Lewis and R.L. Mercer: "Further Results on the Recognition of a Continuously Read Natural Corpus", 1980 IEEE International Conference on ASSP, p.872 (1980-04)

(103) S.E. Levinson: "Maximum Likelihood Parsing of Speech in the Presence of Segmentation Errors", Meeting of the Acoustic Society of America, 67 (S1), S14(A) (1980-04)

(104) D.R. Reddy et al.: "Speech Understanding System - Summary of Results of the Five-Year Research Effort at Carnegie-Mellon Univ.", CMU Tech. Rep. (1977)

(105) W.A. Woods et al.: "Speech Understanding System - Final Technical Progress Report", BBN Tech. Rep. 3438 (1976)

(106) D.E. Walker et al.: "Speech Understanding Research - Final Technical Report", SRI Tech. Rep. (1976)

(107) M.I. Bernstein et al.: "Interactive Systems Research - Final Report", SDC Tech. Rep. (1976)

(108) 鹿野，箱田：＂COSH尺度を用いた会話音声認識システムVoice Q-A System Ⅱの音響処理＂，信学論(D)，Vol.64-D, No.4 (1981-04)

(109) Y. Kobayashi and Y. Niimi: "Word Boundary Detection by Pitch Contours in an Artificial Language", 1980 IEEE International Conference on ASSP, p.900 (1980-04)

(110) 浮田，石川，中川，坂井：＂音声による対話システムにおける発話の確認方法＂，情報処理学会論文誌，Vol.22，No.6，p.589 (1981-11)

(111) 関口，来嘉，重永：＂日本語文章の音声認識システムの改善＂，音学講論，1-7-21 (1980-05)

(112) J.P. Haton and J.M. Pierrel: "Syntactic-Semantic Interpretation of Sentence in the MYRTILLE II Speech Understanding System", 1980 IEEE International Conference on ASSP, p.892 (1980-04)

(113) M. Wagner: "Automatic Labelling of Continuous Speech with a Given Phonetic Transcription Using Dynamic Programming", 1981 IEEE International Conference on ASSP, p.1156 (1981-04)

(114) 佐藤大和："PARCOR-VCV連鎖を用いた音声合成方式"，信学論(D)，Vol.61-D, No.11 (1978)

(115) 箱田，佐藤：＂文音声合成における音調規則＂，信学論(D)，Vol.63-D, No.9 (1980-09)

(116) 中津，好田：＂VCV音節の認識＂,音学講論2-2-16 (1972-10)

(117) Y. Niimi: "A Method for Forming Universal Reference Patterns in an Isolated Word Recognition System", Proc. 4th Int. Joint. Conf. on Pattern Recognition (1978-10)

(118) 中津，好田：＂VCV音節を単位とした単語音声の認識＂，信学技報，PRL73-63 (1973-09)

(119) R. Nakatsu and M. Kohda: "Speech Recognition of Connected Words", Proc. 4th Int. Joint. Conf. on Pattern Recognition (1978-10)

(120) C.S. Myers and L.R. Rabiner: "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition", IEEE Trans. on ASSP, ASSP-29, No.2, p.284 (1981-04)

(121) J.R. Welch and S.C. Oxenberg: "Reduction of Minimum Word-Boundary Gap Lengths in Isolated Word Recognition", 1980 IEEE International Conference on ASSP (1980-04)

(123) 中津，好田："ＶＣＶ音節の認識法の改良"，音学講論，2−4−16 （1975−10）

(124) 中津，好田："ＶＣＶ音節の端点フリーＤＰマッチングを用いた連続単語音声の認識"，信学全大 1210 （1976−03）

(125) 市川 他："連続音韻認識"，音学講論，2−1−14 （1981−10）

(126) 岡 隆一："連続単語認識のためのＤＰアルゴリズムの検討"，音学講論，4−1−22 （1978−05）

(127) M. Kohda, R. Nakatsu and K. Shikano: "Speech Recognition in the Question-Answering System Operated by Conversational Speech", 1976 IEEE International Conference on ASSP (1976-04)

(128) 中津，好田："連続音声のセグメント化と音韻系列への変換"，音声研究会資料 S74−25 （1974−12）

(129) 好田，中津，鹿野："会話音声の認識系の構成"，信学技報，PRL 75−55 （1975−11）

(130) 好田，佐藤，中津，鹿野，箱田："音声による質問回答システム"，信学技報，PRL 75−59，（1976−01）

(131) 鹿野，好田："会話音声の言語処理系の構成"，信学技報PRL75−45 （1975−10）

(132) 中津，好田："会話音声の機械認識における音響処理"，通研実報 Vol.27，№6 （1978−06）

(133) R. Nakatsu and M. Kohda: "An Acoustic Processor in a Conversational Speech Recognition System", Review of the ECL, Vol.26, No.11-12 (1978-12)

(134) 中津．好田："Top−down的処理を付加した連続音声の音響処理の検討"信学技報 PRL76−34 （1976−09）

(135) 好田，中津，鹿野："オンライン会話音声認識系の構成"，音声研究会資料S76−28 （1976−11）

(136) 中津，鹿野，伊藤，好田："音声によるオンライン質問回答システム"，音声研究会資料S77−14 （1970−06）

(137) 鹿野，中津，好田："オンライン会話音声認識系の性能"，音声研究会資料，S77−13 （1977−06）

(138) 中津，好田："会話音声認識における標準パターンの検討"，音学講論，3−1−16 （1978−10）

(139) 中津良平："単語音声認識における距離尺度の検討"，信学全大 1292 （1979−03）

(140) 長島，中津：＂音韻単位の標準パターンを用いた実時間単語音声認識装置＂，音声研究会資料 S78−22 (1978−22)
(141) 長島，中津：＂音韻標準パターン学習装置＂，信学会部門別全国大会 55(1979−10)
(142) 中津，長島：＂連続単語音声認識方法の検討＂，信学論(D)，Vol.64−D，12 (1981−12)
(143) 中津，長島，嵯峨山：＂連続単語音声認識方式＂，通研実報，Vol.30，No.10 (1981−10)
(144) R. Nakatsu: "A Speech Recognition Machine for Connected Words", 1980 International Conference on ASSP (1980-04)
(145) 中津，長島：＂連続単語音声認識装置＂，音声研究会資料，S79−54 (1979−12)
(146) 中津，好田：＂オンライン会話音声認識における音響処理部の構成＂，音学講論，2−4−7 (1976−10)
(147) 中津，好田：＂オンライン会話音声認識系におけるＶＣＶ音節認識の高速化＂，信学全大，1242 (1977−10)
(148) 鹿野清宏：＂音声ラティスの評価システム＂，信学論(A)，Vol.63−A，No.3 (1980−03)
(149) S.E.G. Ohman: "Coarticulation in VCV Utterance: Spectrographic Measurements", Journal of the Acoustic Society of America, Vol.39, No.1 (1966)
(150) 長島，好田：＂音韻の標準パターンを用いた単語音声の認識＂，音学講論，2−4−3 (1976−05)
(151) 好田，板倉，斎藤：＂音声情報高速処理装置＂，通研実報，Vol.25，No.4 (1976−04)

付　　録

1. ＶＣＶ音節認識の confusion matrix
2. 会話音声認識システム（第１次）の発声リスト
3. 会話音声認識システム（第１次）における音声データのソナグラム
4. 会話音声認識システム（第２次）の言語情報
5. 会話音声認識システム（第２次）の発声リスト
6. 単語音声認識実験における単語辞書

348項欠

付録 1. VCV音節認識実験の confusion matrix

（単独発声のＶＣＶ音節）

表A 1.1　実 験 1 の 結 果

a.

入力＼出力	a	i	u	e	o	認識率(%)
a	68				2	97.1
i		69	1			98.6
u			70			100
e		23		47		67.1
o			3		67	95.7
					計	91.7%

b.

入力＼出力	m	n	b	d	g	r	z	認識率(%)	
m	16	7		1		1		64.0	94.0
n	4	20					1	80.0	
b			24	1				96.0	85.7
d	1		18	3		1	2	12.0	
g	5		15	1	2	2		8.0	
r	1	2	2	2		18		72.0	
z	1	1	6	1		10	6	24.0	
							計	50.9%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	14	4		4		3
n	17	1	2		1	4
b	20					5
d	2	6	7			10
g	2		1	1	1	20
r	17	1	1	1	1	4
z	5	7	3	3	1	6
計	77	19	14	9	4	52
%	44.0	10.9	8.0	5.1	2.3	29.7

表A 1.2　実験 2 の結果

a.

入力＼出力	a	i	u	e	o	認識率(%)
a	70					100
i		70				100
u			70			100
e				70		100
o					70	100
					計	100%

b.

入力＼出力	m	n	b	d	g	r	z	認識率(%)	
m	15	5	1			1	3	60.0	88.0
n	4	20					1	80.0	
b		1	9	5	2		8	36.0	82.7
d			7	13	1		4	52.0	
g			6	1	18			72.0	
r			2	1		15	7	60.0	
z			5	1		4	15	60.0	
							計	60.0%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	15	5	1	3		1
n	20	4		1		
b	9	11	5			
d	13	6	4	2		
g	18	3	2	2		
r	15	8	2			
z	15	9				1
計	105	46	14	8		2
%	60.0	26.3	8.0	4.6		1.1

表A 1.3　実験 3 の結果

a．

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		70				100
u			70			100
e				70		100
o					70	100
					計	100%

b．

入力＼出力	m	n	b	d	g	r	z	認識率	
m	17	7	1					68.0	96.0
n	6	18		1				72.0	
b			11	12	2			44.0	96.0
d			2	23				92.0	
g	3		3	10	9			36.0	
r	1	1	1	8		10	4	40.0	
z	1	1	3	9			11	44.0	
							計	56.6%	

c．

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	17	5	2	1		
n	18	7				
b	11	9	4	1		
d	23	2				
g	9	5	5	3	2	1
r	10	3	2	2	3	5
z	11	6	6		1	1
計	99	37	19	7	6	7
%	56.6	21.1	10.9	4.0	3.4	4.0

表A 1.4　　実験 4 の結果

a.

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		70				100
u			70			100
e				70		100
o					70	100
					計	100%

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	25							100	100
n	2	23						92.0	
b			25					100	98.7
d				24			1	96.0	
g			1		24			96.0	
r						25		100	
z							25	100	
							計	97.7%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	25					
n	23	2				
b	25					
d	24	1				
g	24	1				
r	25					
z	25					
計	171	4				
%	97.7	2.3				

表A 1.5　実験 5 の結果

a.

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		70				100
u			70			100
e				70		100
o					70	100
					計	100%

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	23	1	1					92.0	98.0
n		25						100	
b			25					100	98.7
d			2	23				92.0	
g			2		22	1		88.0	
r						24	1	96.0	
z						1	24	96.0	
							計	94.9%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	23	2				
n	25					
b	25					
d	23	2				
g	22	3				
r	24	1				
z	24			1		
計	166	8		1		
%	94.9	4.6		0.6		

表A 1.6　　実験 6 の結果

a.

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		70				100
u			70			100
e				70		100
o					70	100
					計	100%

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	24	1						96.0	100
n	3	22						88.0	
b			23	2				92.0	98.7
d				25				100	
g			3	1	20		1	80.0	
r	2		2			20	1	80.0	
z			1				24	96.0	
							計	90.3%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	24	1				
n	22	3				
b	23	2				
d	25					
g	20	4	1			
r	20	2	2	1		
z	24	1				
計	158	13	2	1		
%	90.3	7.4	1.7	0.6		

表A 1.7　実験 7 の結果

a.

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		69	1			98.6
u			69		1	98.6
e				70		100
o			2		68	97.1
					計	98.9%

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	10	8	3		1	1	2	40.0	78.0
n	3	18	1			2	1	72.0	
b		1	10	3	4		7	40.0	72.0
d		2	5	9	3	2	4	36.0	
g			3	1	16	1	4	64.0	
r		1	2	3	2	10	7	40.0	
z		1	2	2	1	5	14	56.0	
							計	49.7%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	10	8	2	3	1	1
n	18	3	1	1	1	1
b	10	5	1	2	3	4
d	9	7	4	5		
g	16	3	1		1	4
r	10	5	6	2	1	1
z	14	1	2	5	2	1
計	87	32	17	18	9	12
%	49.7	18.3	9.7	10.3	5.1	6.9

表A 1.8　実験 8 の結果

a.

入力＼出力	a	i	u	e	o	認識率
a	70					100
i		70				100
u		1	62		7	88.6
e				70		100
o	2		2		66	94.3
					計	96.6 %

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	15	5	4	1				60.0	84.0
n	7	15	1	1			1	60.0	
b	2	2	17	3		1		68.0	82.0
d		1	12	6	1	2	3	24.0	
g			12	4	7	1	1	28.0	
r	3		9	2		11		44.0	
z	1	1	9	2		6	6	24.0	
							計	44.0%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	14	6	3	1		1
n	12	9	2		2	
b	17	4	1	2	1	
d	6	10	2	3	3	1
g	6	6	5	1	3	4
r	10	7	6	1		1
z	6	2	3	2	2	10
計	71	44	22	10	11	17
%	40.6	25.1	12.6	5.7	6.3	9.7

表A1.9　実験9の結果

a

入力＼出力	a	i	u	e	o	認識率
a	69				1	98.6
i		70				100
u			70			100
e				70		100
o					70	100
					計	99.7%

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	14	8			2	1		56.0	88.0
n	1	21		1	1	1		84.0	
b	1		19	4		1		76.0	88.0
d			3	14	1	2	5	56.0	
g			1	1	23			92.0	
r		3	1	8		13		52.0	
z		1	1	6			17	68.0	
							計	69.0%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	14	8	1	2		
n	21	2	2			
b	19	4	1		1	
d	14	8	1	1	1	
g	23	2				
r	13	5	4	3		
z	17	5	2	1		
計	121	34	11	7	2	
%	69.1	19.4	6.3	4.0	1.1	

表A 1.10　実験 10 の結果

a.

出/入	a	i	u	e	o	認識率
a	120					100
i		120				100
u			120			100
e				120		100
o					120	100
					計	100%

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	20	4	1										80.0	96.0
n	2	22		1									88.0	
b	1	1	13	2			1	7					52.0	62.7
d			1	13		1	5	3	2				52.0	
g			1	3	14	1	2	1	1	2			56.0	
r				1		18	5	1					72.0	
z				1			23		1				92.0	
p			8	1		2		11	3				44.0	74.7
t				1			1	2	18		3		72.0	
k					1	1	1	1	1	20			80.0	
s				1		2	2				19	1	76.0	
h										1		24	96.0	
												計	71.7%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	20	5				
n	22	3				
b	13	5	3	2		2
d	13	8	1	2		1
g	14	3	2	2	2	2
r	18	5	2			
z	23	2				
p	11	8	3	2		1
t	18	2	2	2		1
k	20	2	1	1		1
s	19	2	3		1	
h	24	1				
計	215	46	17	11	3	8
%	71.7	15.3	5.7	3.7	1.0	2.7

表A 1.11　実験 11 の結果

a.

入＼出	a	i	u	e	o	認識率
a	120					100
i		120				100
u		1	118	1		98.3
e				120		100
o			1		119	99.2
						99.5%

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	19	3	1	1				1					76.0	86.0
n	5	16	1	1		1	1						64.0	
b	1	1	10	3			1	7	2				40.0	61.3
d		4	3	13		1	1	1	2				52.0	
g	1	1	1	5	11	2			1	3			44.0	
r	1	1	1	5		11	6						44.0	
z			1	5			18		1				72.0	
p			10	2			1	8	2	1	1		32.0	49.3
t			2	3	1	1	2	3	10	1	2		40.0	
k		1	2	1	5			1	2	9	1	3	36.0	
s			1	3			1	1	1		16	1	64.0	
h								1				24	96.0	
												計	55.0%	

c.

＼	1　位	2　位	3　位	4　位	5　位	6位以下
m	19	5		1		
n	16	6	1	2		
b	9	5	4	3	1	3
d	13	4	1	1	3	3
g	11	2	4	2	1	5
r	11	8	2	1	1	2
z	18	4	1	1	1	
p	8	4	6	3		4
t	10	6	3	1	2	3
k	9	6	2	3		5
s	15	1	1	1		7
h	24					1
計	163	51	25	19	9	33
%	54.3	17.0	8.3	6.3	3.0	11.0

表A 1.12　実験 12 の結果

a.

入＼出	a	i	u	e	o	認識率
a	119			1		99.2
i		118	1	1		98.3
u		2	112		6	93.3
e		1	1	118		98.3
o			4		116	96.7
					計	97.2%

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	14	8		1				2					56.0	78.0
n	5	12	2	3		1	1				1		48.0	
b	1	2	2	7			2	6	3	2			8.0	37.3
d	1	1		7		2	6	5	3				28.0	
g			1	1	10	2		3	1	5		2	40.0	
r		2	2			14	4	2		1			56.0	
z				3		2	18		1	1			72.0	
p		1	5	5	1		1	8	2		1	1	32.0	46.7
t		2		2	2		2	2	12		2	1	48.0	
k		1	3	1	6			1	4	6		3	24.0	
s	1			1		2	4		1	1	14	1	56.0	
h	2	1		1				1				20	80.0	
												計	45.7%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	12	5	1		2	5
n	12	5	1		1	6
b	2	4	3	3	3	10
d	5	5	2	1	2	10
g	10	4	1	2	1	7
r	13	3	2		1	6
z	18	6				1
p	8	4	3	2	1	7
t	12	2	2	3	1	5
k	6	6	2	1	1	9
s	14	1	3	2	1	4
h	19	3	2	1		
計	131	48	22	15	14	70
%	43.7	16.0	7.3	5.0	4.7	233

表A 1.13　実験 13 の結果

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	21	2	1					1					84.0	94.0
n	1	23		1									92.0	
b			21	2		1	1						84.0	90.7
d			1	20	1	1	2						80.0	
g			1		22					2			88.0	
r		2		6		17							68.0	
z		3		4			16		1		1		64.0	
p			12	3				10					40.0	66.7
t				3	1				19	1		1	76.0	
k					4					21			84.0	
s				1	1			1			20	2	80.0	
h												25	100	
												計	78.3%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	21	1	1	1	1	
n	23	2				
b	21	3				1
d	20	3		1		1
g	22	3				
r	17	6	1	1		
z	16	8	1			
p	10	6	5	2	1	1
t	19	3	1	1	1	
k	21	3			1	
s	20	3	1			1
h	25					
計	235	41	10	6	4	4
%	78.3	13.7	3.3	2.0	1.3	1.3

表A 1.14　実験 14 の結果

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	22	3											88.0	
n	1	22		1		1							88.0	
b			23	1			1						92.0	
d			2	20		2	1						80.0	94.7
g				1	24								96.0	
r		2	1	3		19							76.0	
z		1	1	4			18		1				72.0	
p			14	3				8					32.0	
t		1	1	1	1			1	19	1			76.0	68.0
k					3					22			88.0	
s							1		1		22	1	88.0	
h												25	100	
												計	81.3%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	22	2	1			
n	22	3				
b	23	1		1		
d	20	4	1			
g	24		1			
r	19	3	2	1		
z	18	7				
p	8	11	4	2		
t	19	3		1		2
k	22	2				1
s	22	2	1			
h	25					
計	224	38	10	5		3
%	81.3	12.7	3.3	1.7		1.0

表A 1.15　実験 15 の結果

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	22	3											88.0	96.0
n	1	22		1		1							88.0	
b			23	1			1						92.0	94.7
d			2	20		2	1						80.0	
g				1	24								96.0	
r		2	1	3		19							76.0	
z		1	1	4			18		1				72.0	
p								25					100	97.3
t								2	20	2	1		80.0	
k					1					24			96.0	
s									2		22	1	88.0	
h												25	100	
												計	88.0%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	22	2	1			
n	22	3				
b	23		1		1	
d	20	4	1			
g	24		1			
r	19	3	2	1		
z	18	7				
p	25					
t	20	4	1			
k	24	1				
s	22	3				
h	25					
計	264	27	7	1	1	
%	88.0	9.0	2.3	0.3	0.3	

表A 1.16　実験 16 の結果

b

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	23	2											92.0	100
n	2	23											92.0	
b	2		22	1									88.0	96.0
d				24			1						96.0	
g					25								100	
r				2		21	2						84.0	
z			2	2			21						84.0	
p								24	1				96.0	100
t								4	21				84.0	
k								3		22			88.0	
s									2		23		92.0	
h										2		24	96.0	
												計	91.0%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	23	2				
n	23	2				
b	22	3				
d	24	1				
g	25					
r	21	3	1			
z	21	4				
p	24	1				
t	21	3	1			
k	22	3				
s	23	2				
h	24	1				
計	273	25	2			
%	91.0	8.3	0.6			

表A 1.17　実験 17 の結果

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	20	3	1					1					80.0	88.0
n		21					2	1	1				84.0	
b			20		2		1	2					80.0	84.0
d			2	17		1	5						68.0	
g	2		1	1	20			1					80.0	
r	1			4		19	1						76.0	
z		2	1				22						88.0	
p	1	1	5		1		1	14			1	1	56.0	57.3
t	2			6	1		3	3	8		2		32.0	
k	1	1		1	2		1		3	15	1		60.0	
s	1	2		1			1	2	3		13	2	52.0	
h					1					2		22	88.0	
												計	70.3%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	20				2	3
n	21	1		2		1
b	20	3			1	1
d	17	5	2			1
g	20	2	2			1
r	19	3	1	2		
z	22	2				1
p	14	3	3	1	1	3
t	8	8	1	1	2	5
k	15	3			2	5
s	13	2	3	1		6
h	22	2	1			
計	211	34	13	7	8	27
%	70.3	11.3	4,3	2.3	2.7	9.0

表A 1.18　実験 18 の結果

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	21	1	2		1								84.0	
n	3	22											88.0	94.0
b	2		18	1			1	3					72.0	
d			1	18		2	4						72.0	80.0
g	1				22			1		1			88.0	
r		2		3		19	1						76.0	
z		2		2			20				1		80.0	
p	1		6	2				16					64.0	
t			1		2		2	2	17	1			68.0	78.7
k	1						1			23			92.0	
s		1							1		20	3	80.0	
h												25	100	
												計	80.3%	

	1　位	2　位	3　位	4　位	5　位	6位以下
m	21	2	2			
n	22	3				
b	18	6	1			
d	18	5	1			1
g	22	1	1	1		
r	19	3		2	1	
z	20	3	1		1	
p	16	5	3		1	
t	17	4	1		3	
k	23		1			1
s	20	3		1		1
h	25					
計	241	35	11	4	6	3
%	80.3	11.7	3.7	1.3	2.0	1.0

表A 1.19　実験 19 の結果

b.

入力＼出力	m	n	b	d	g	r	z	認識率	
m	22		2		1			88.0	92.0
n	2	22		1				88.0	
b			24	1				96.0	98.7
d			2	22		1		88.0	
g			2		23			92.0	
r		2	1	4		18		72.0	
z							25	100	
							計	89.1%	

c.

子音＼順位	1　位	2　位	3　位	4　位	5　位	6位以下
m	22	3				
n	22	2	1			
b	24	1				
d	22	2	1			
g	23	2				
r	18	4	2	1		
z	25					
計	156	14	4	1		
%	89.1	8.0	2.3	0.6		

表A1.20　実験20の結果

b.

	m	n	b	d	g	r	z	p	t	k	s	h	認識率	
m	21	3	1										84.0	92.0
n		22		1		1			1				88.0	
b			22	2			1						88.0	94.7
d			3	19		1	2						76.0	
g				1	24								96.0	
r		2	1	3		19							76.0	
z		1	1	3			20						80.0	
p								25					100.	100
t								1	22	2			88.0	
k								1		24			96.0	
s									2		22	1	88.0	
h												25	100	
												計	88.3%	

c.

	1　位	2　位	3　位	4　位	5　位	6位以下
m	21	3	1			
n	22	2				1
b	22	2		1		
d	19	5	1			
g	24		1			
r	19	5		1		
z	20	4	1			
p	25					
t	22	2		1		
k	24	1				
s	22	2	1			
h	25					
計	265	26	5	3		1
%	88.3	8.7	1.7	1.0		0.3

付録 2. 会話音声認識システム（第１次）の発声リスト

(1) 東京から，新大阪までの，８時４５分発の，ひかり６１号で，グリーン券を，４枚，予約します。

(2) 新横浜発，８時９分の，指定席を，静岡まで，お願い致します。

(3) １０時３分発の，ひかりで，名古屋から，京都まで，２枚，予約します。

(4) 小田原を，９時３分発で，こだま号の，グリーン券を，米原まで，お願いします。

(5) 東京，８時３０分発の，ひかり５９号の，指定券を，３枚，予約します。

(6) 岐阜羽島までで，熱海発，９時４５分の，こだま１１９号の，普通席を，６枚，予約します。

(7) 名古屋発，１１時３分発の，ひかり２７号で，新大阪ゆきの，指定券は，ありますか。

(8) 三島を，１０時２５分発の，こだま号の，豊橋ゆきの，グリーン券を，予約します。

(9) ９時３０分発，東京駅で，ひかり号の，名古屋ゆきの，グリーン券を，９枚。

(10) こだま１２３号で，静岡を，１１時２１分発の，米原いきの，指定券を，５枚。

(11) １０時，東京発の，ひかり２９号の，新大阪いきの，グリーンを，１枚，とります。

(12) 名古屋駅，１２時１８分の，ひかり３号の，指定券を，京都まで，７枚。

(13) こだまで，浜松発，１２時１９分の，新大阪への，指定券を，２枚。

(14) 名古屋から，１２時３３分発の，ひかり６７号で，新大阪まで，グリーン券は，ありますか。

(15) １３時１０分発の，こだまで，豊橋から，米原までの，グリーン席は，ありますか。

(16) ひかり３３号で，１１時発で，東京からの，名古屋まで，グリーン券を，５枚。

(17) 小田原，１２時３分発の，こだま１３３号の，三島ゆき，指定券は，ありますか。

(18) １１時３０分，東京発の，新大阪までの，ひかりの，普通席は，ありますか。

(19) 名古屋を，１３時４８分発の，ひかりで，京都までの，指定券を，４枚，お願い致します。

(20) 新横浜から，１２時９分の，こだま１３５号の，グリーン席を，６人，浜松まで，お願い致します。

(21) １２時，東京発の，指定券を，名古屋まで。

(22) １３時１５分発，小田原で，米原までの，指定席を，５人。

(23) ひかり７５号で，名古屋から，新大阪まで。

(24) １３時５５分，三島発で，米原まで，グリーンを，３枚。

(25) １３時０分発，ひかり号で，東京から，新大阪まで，６枚。

(26) 岐阜羽島発，１６時２１分で，京都までの，グリーン席を，２枚。

(27) 東京発，１３時３０分の，ひかり号の，指定券を，名古屋まで。

(28) 静岡から，岐阜羽島までの，こだま１４５号の，指定券を，７枚。

(29) こだま１１６号で，京都から，豊橋まで。

(30) 新大阪，８時２５分発で，東京までの，グリーン券を，３枚。

(31) 米原より，９時３１分発の，小田原への，指定。

(32) ひかり２４号で，京都から，東京まで，４枚。

(33) １０時２２分の，岐阜羽島から，三島への，グリーンを，４人。

(34) 名古屋駅からで，１０時３３分発の，ひかり５８号で，東京まで，８枚。

(35) こだま１２８号で，１１時３４分発で，東京まで。

(36) ひかり２６号の，指定券で，新大阪から，名古屋まで。

(37) １１時１分発の，こだま１３０号で，新横浜まで。

(38) 京都駅，１０時４４分発で，東京まで，２枚。

(39) こだま１３４号で，米原から，浜松まで。

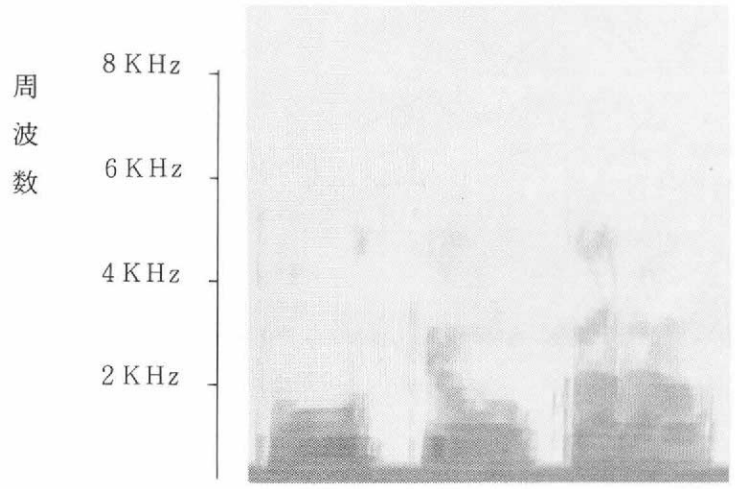
(40) ひかり２８号で，新大阪より，東京まで。

(41) 静岡，１３時５３分発で，新横浜へ，普通席，５枚。

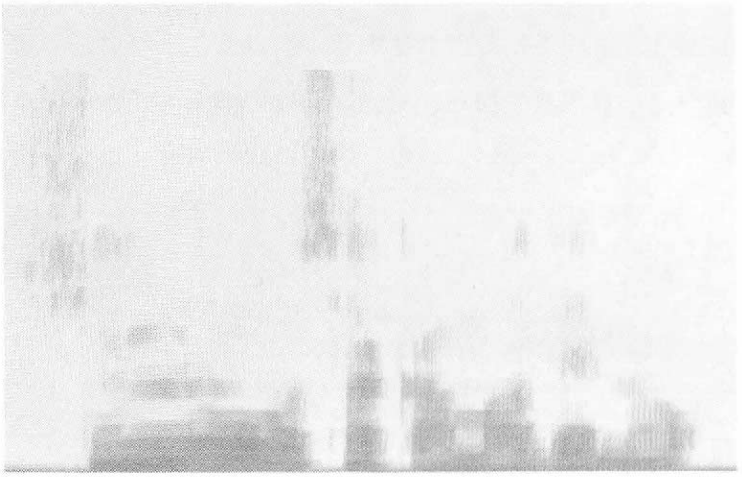
(42) 京都駅より，名古屋まで，ひかり６４号で，４枚。

付録 3. 会話音声認識システム（第１次）における音声データのソナグラム
（文章(1)～(10)の全文節およびその他の重要な文節）

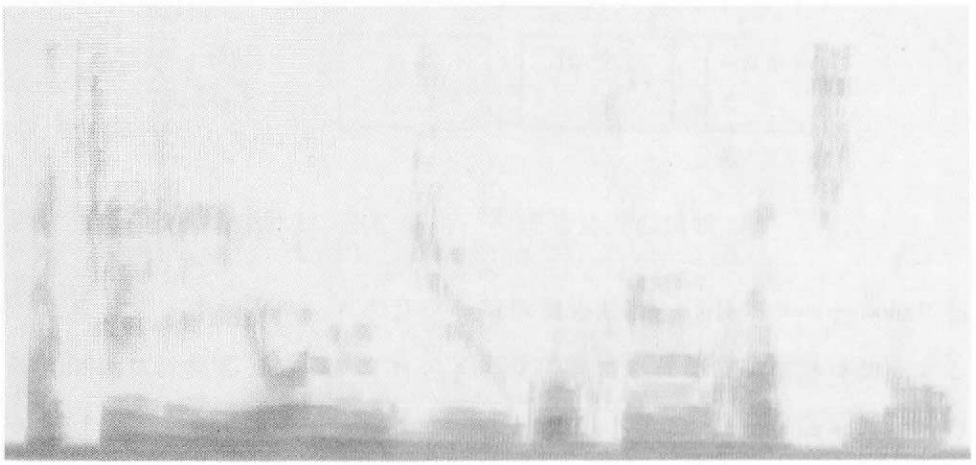
(1) 東京から



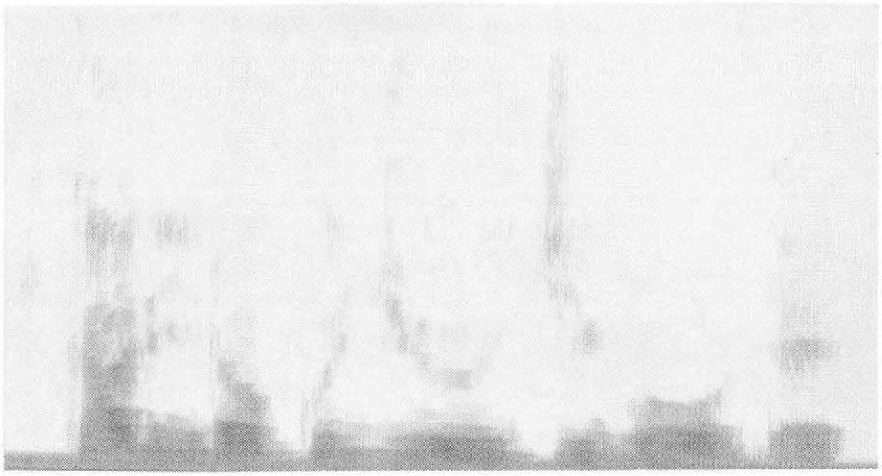
(1) 新大阪までの



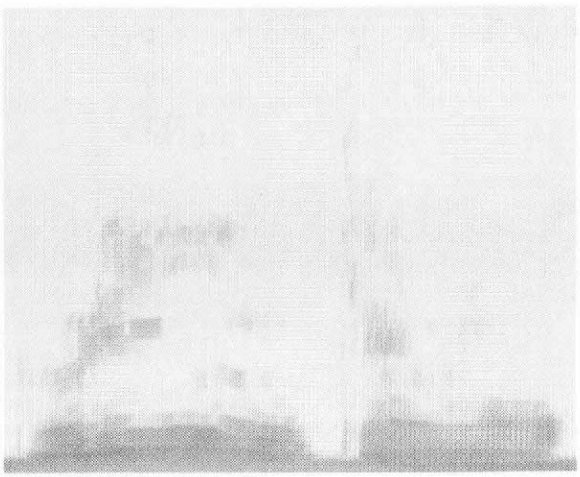
(1) ８時４５分発の



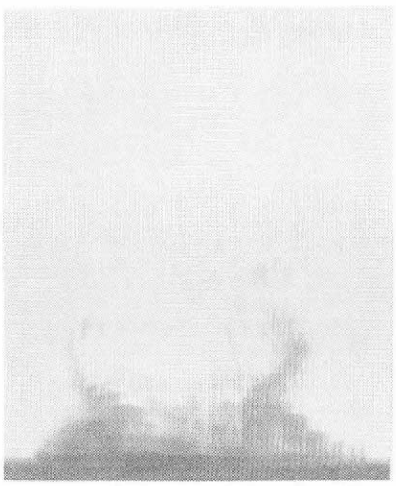
(1)　ひかり６１号で



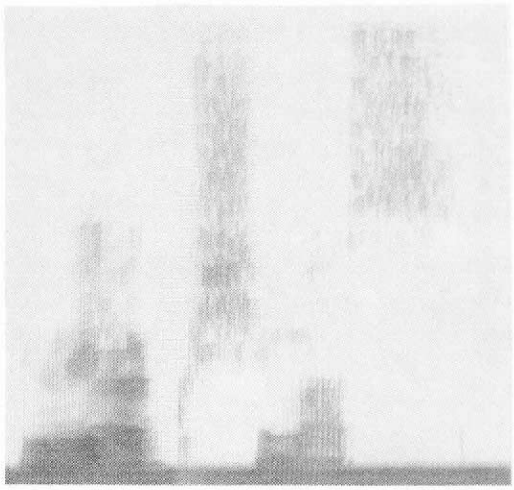
(1)　グリーン券を



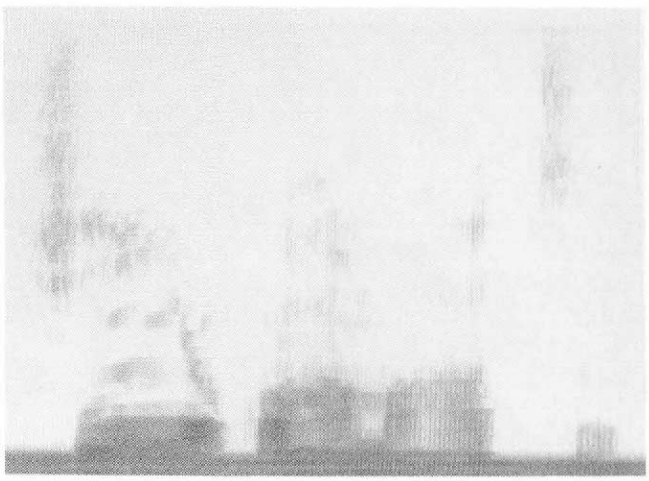
(1)　４枚



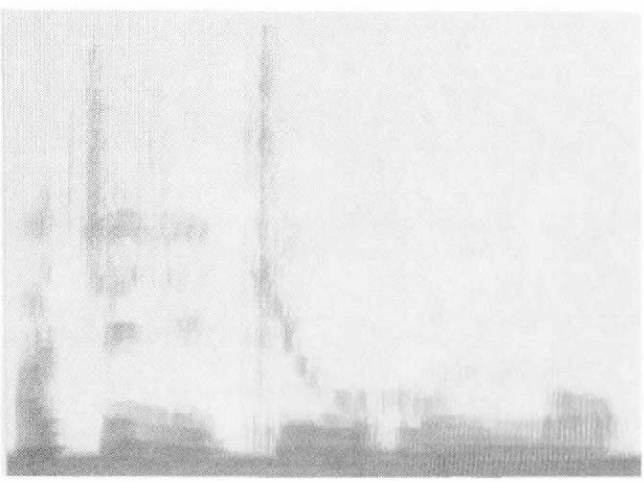
(1)　予約します



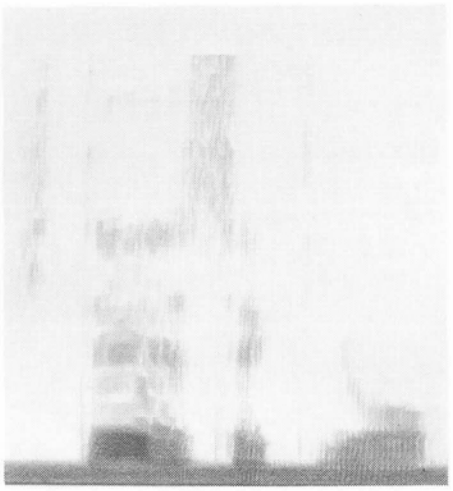
(2)　新横浜発



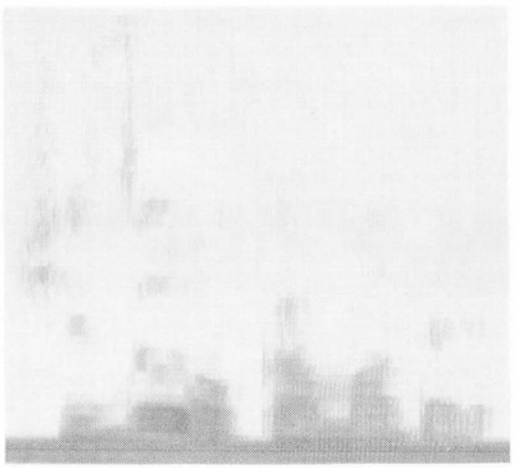
(2)　8時9分の



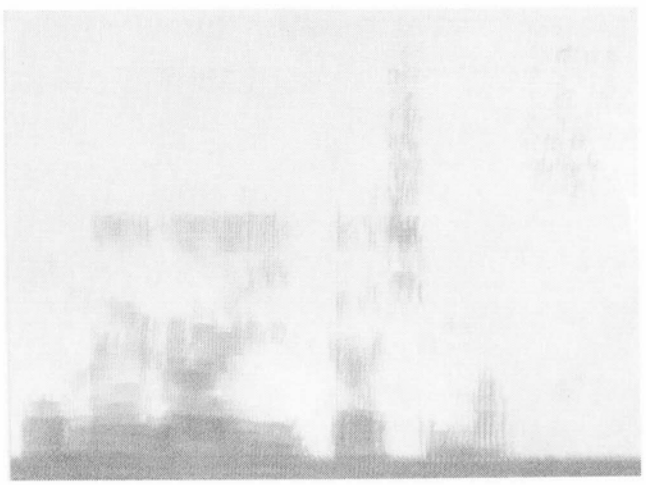
(2)　指定席を



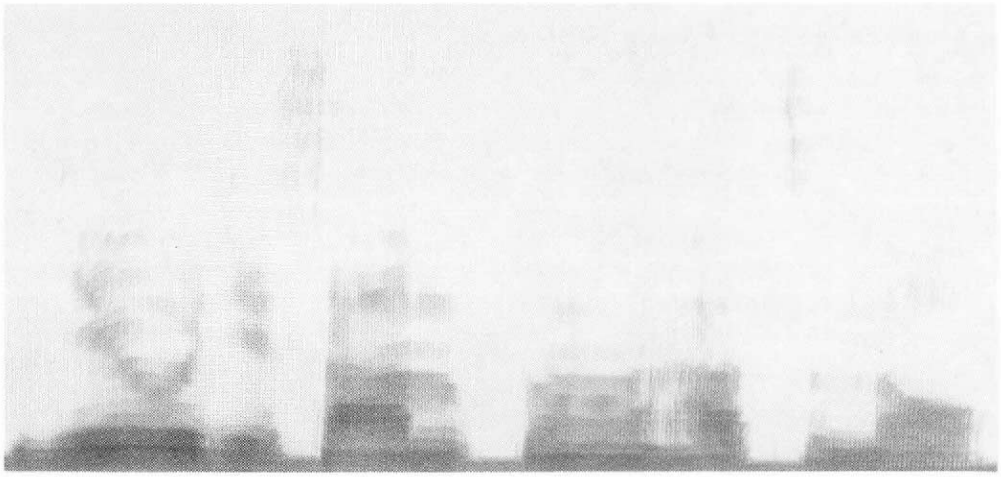
(2)　静岡まで



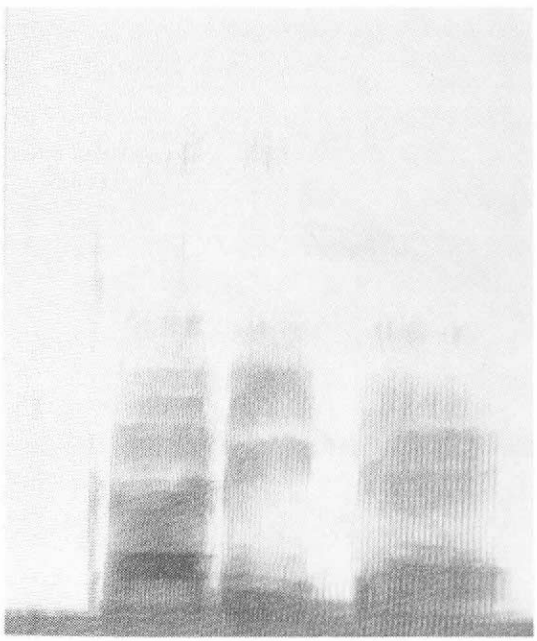
(2)　お願い致します



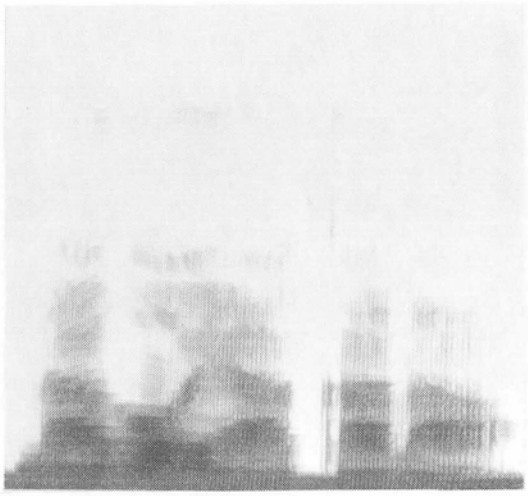
(3)　10時3分発の



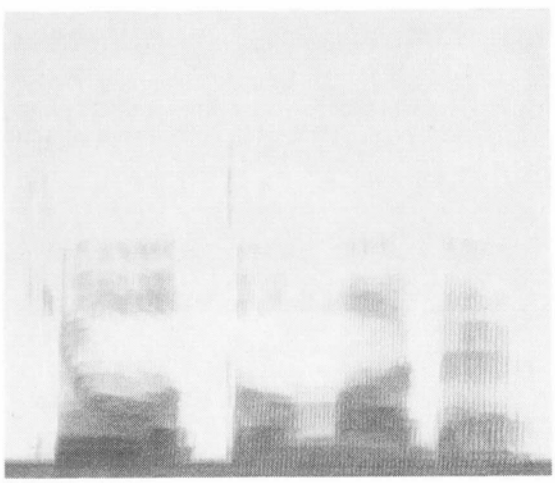
(3)　ひかりで



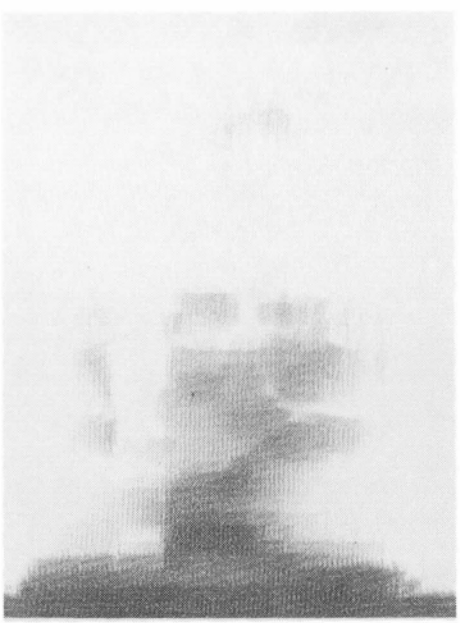
(3)　名古屋から



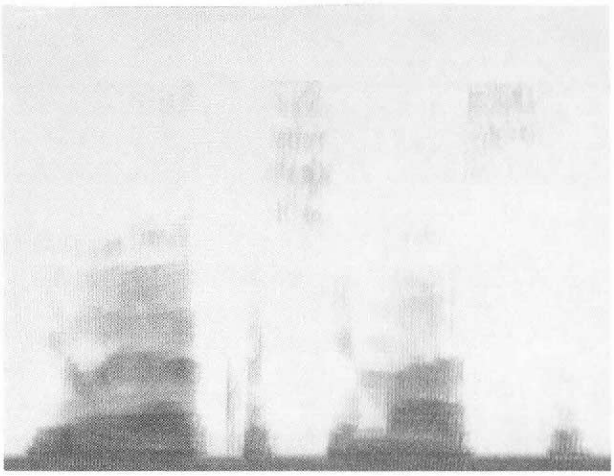
(3)　京都まで



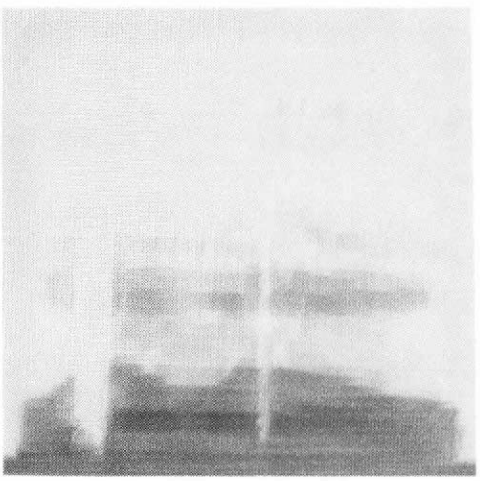
(3)　2 枚



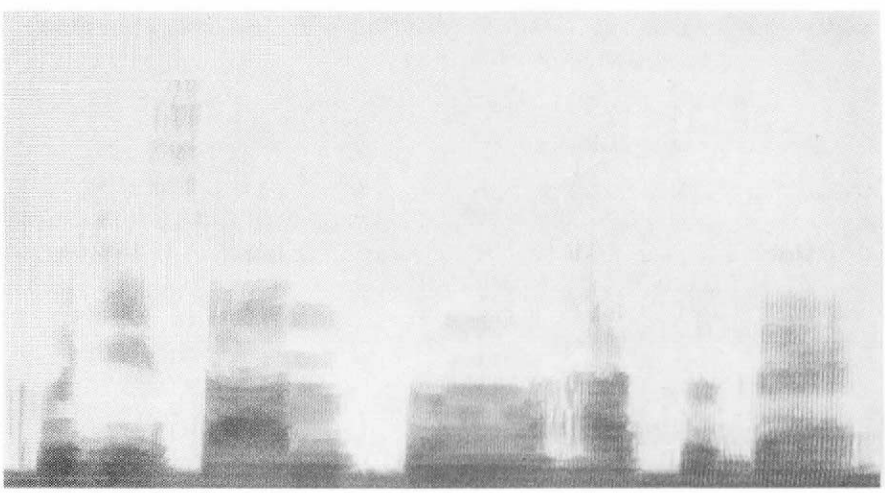
(3)　予約します



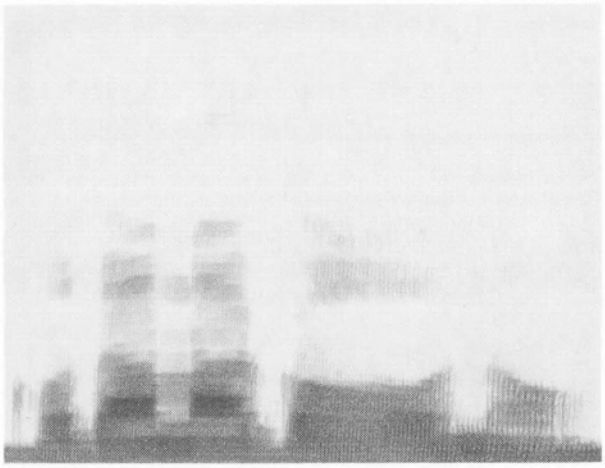
(4)　小田原を



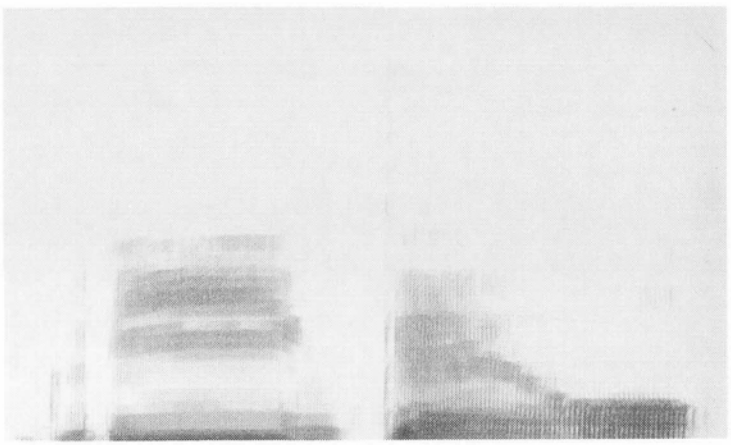
(4)　9時3分発で



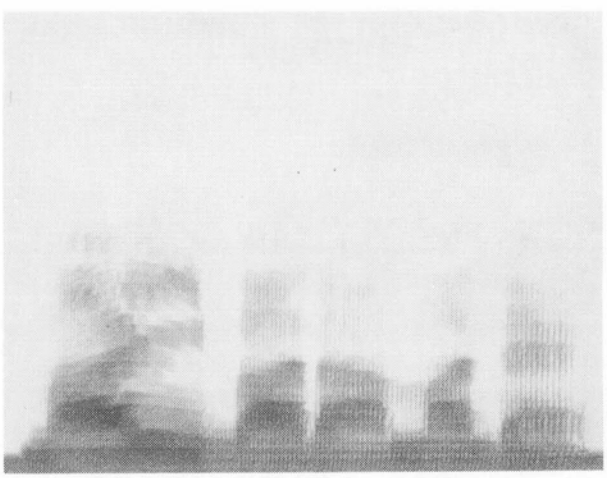
(4)　こだま号の



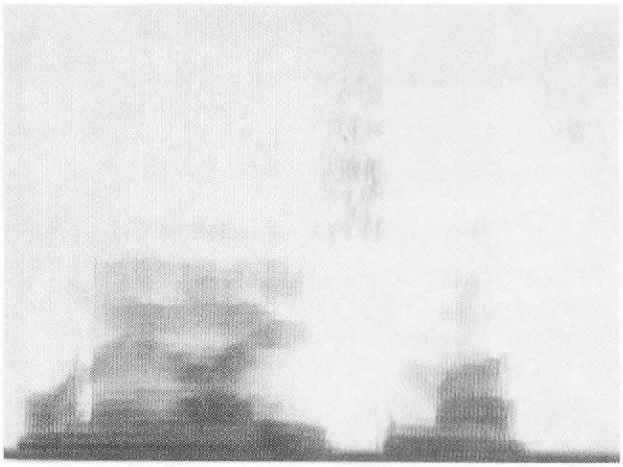
(4)　グリーン券を



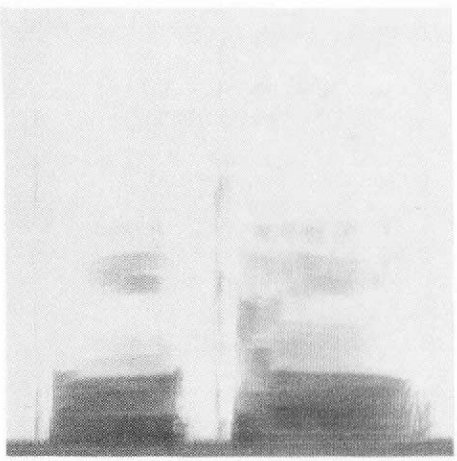
(4)　米原まで



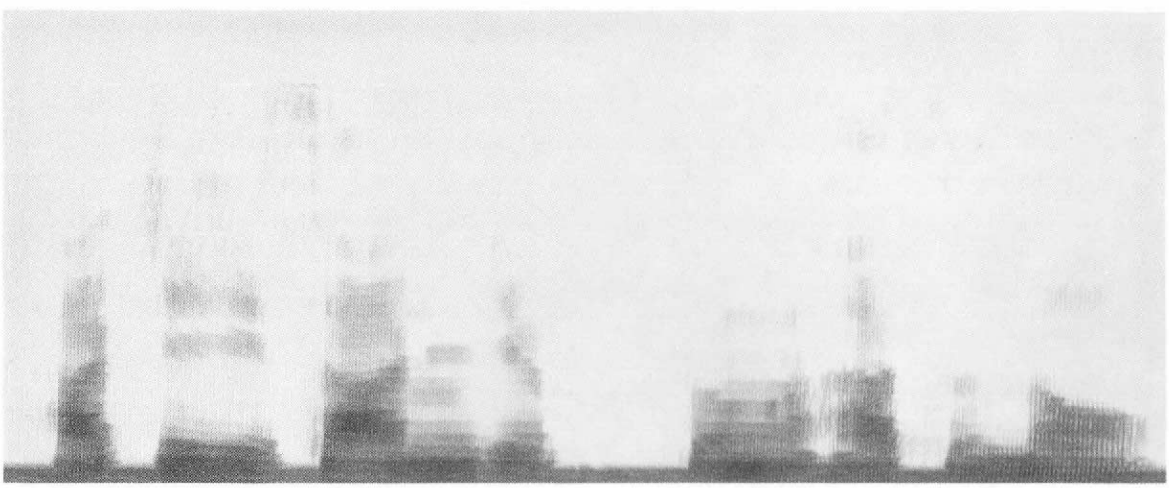
(4)　お願いします



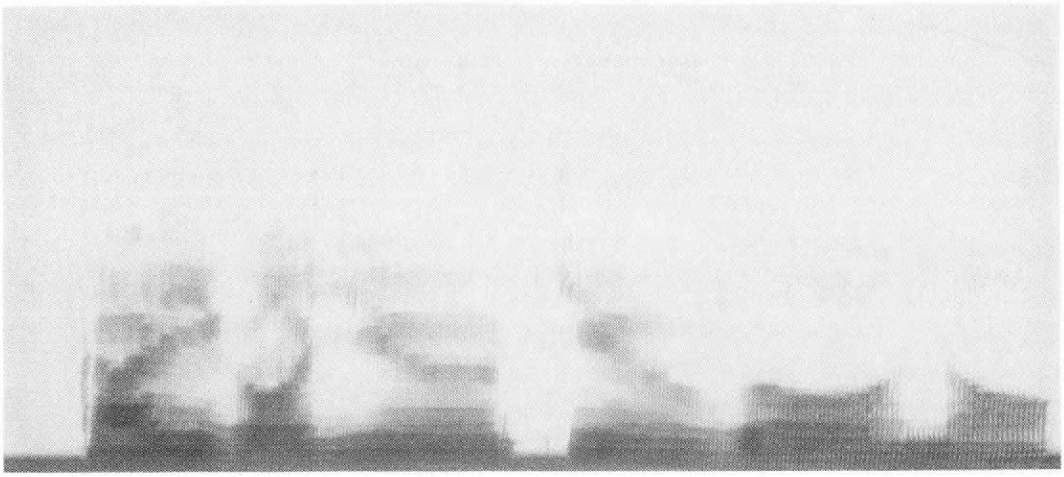
(5)　東京



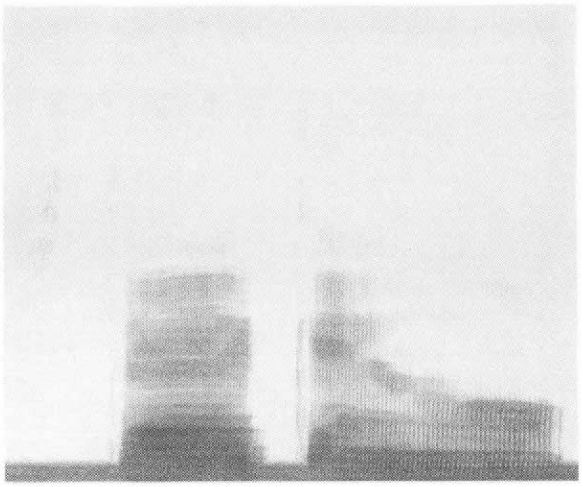
(5)　8 時 30 分発の



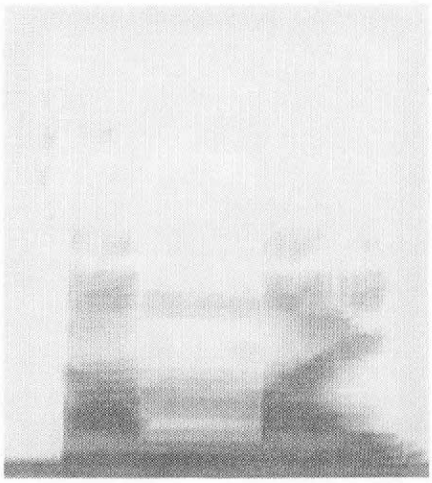
(5)　ひかり５９号の



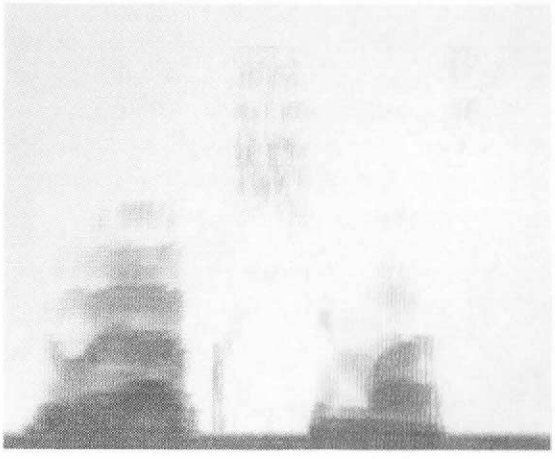
(5)　指定券を



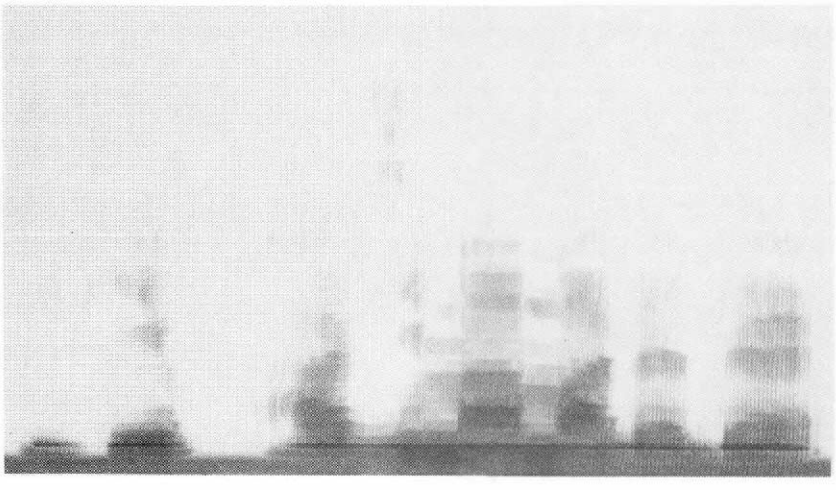
(5)　３枚



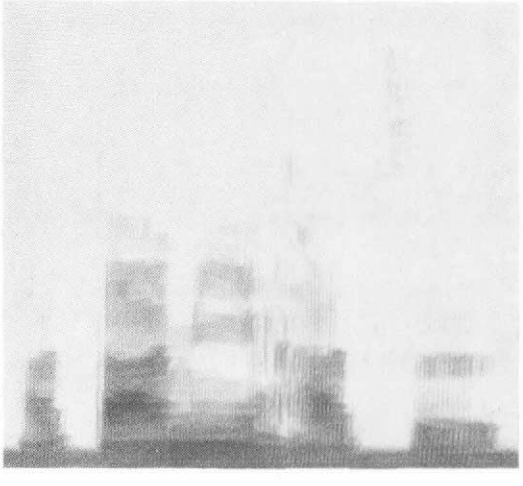
(5)　予約します



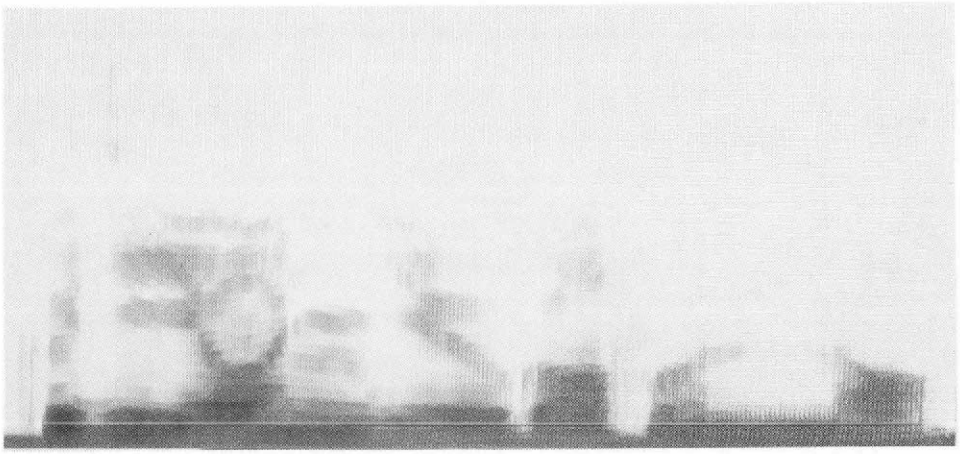
(6)　岐阜羽島までで



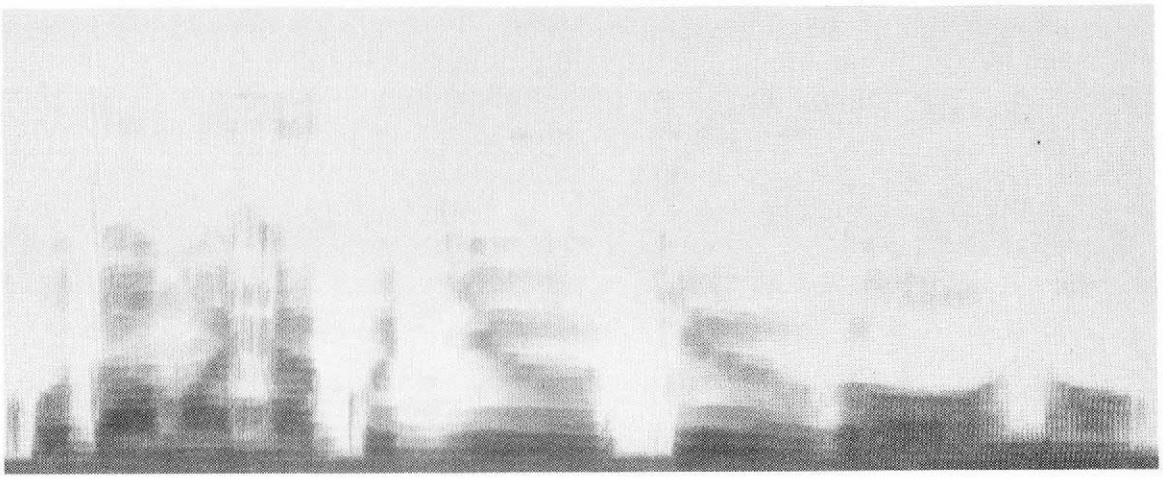
(6)　熱海発



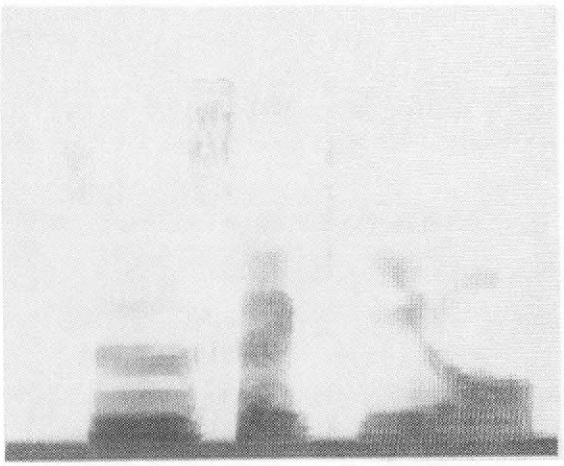
(6)　9時45分の



(6)　こだま119号の

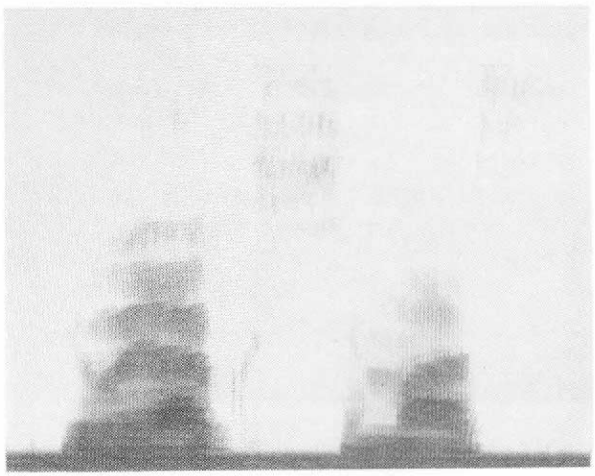


(6)　普通席を

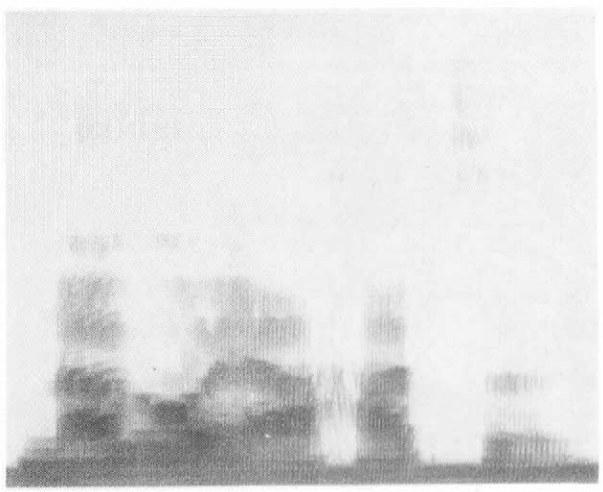


(6)　6 枚

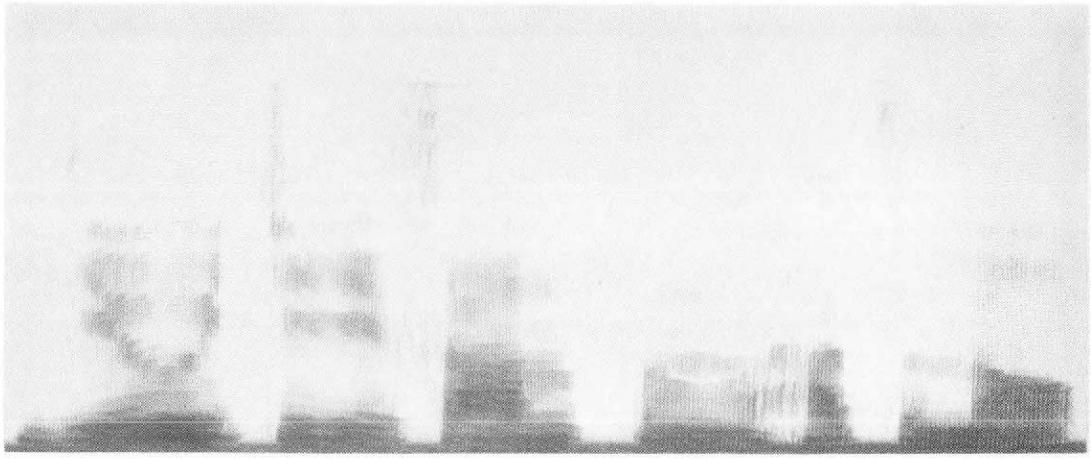
(6)　予約します



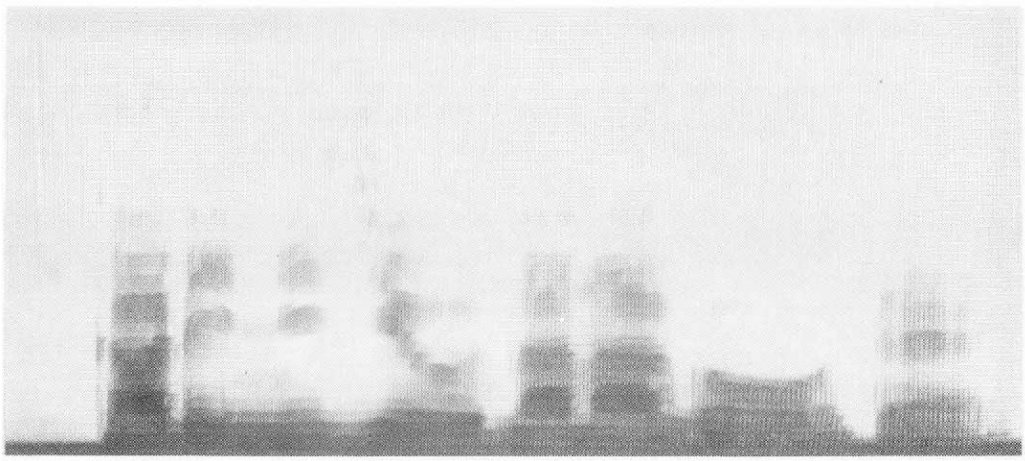
(7)　名古屋発



(7)　11時3分発の

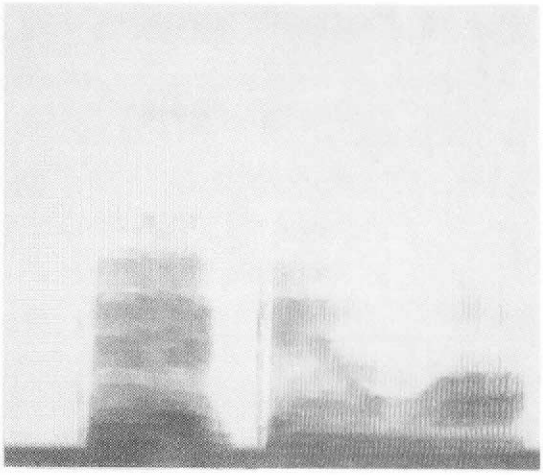


(7)　ひかり27号で

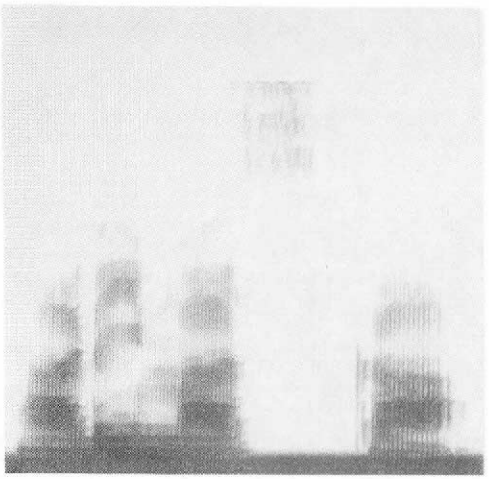


(7)　新大阪ゆきの

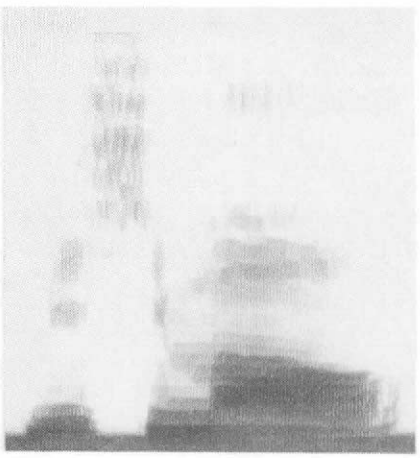
(7)　指定券は



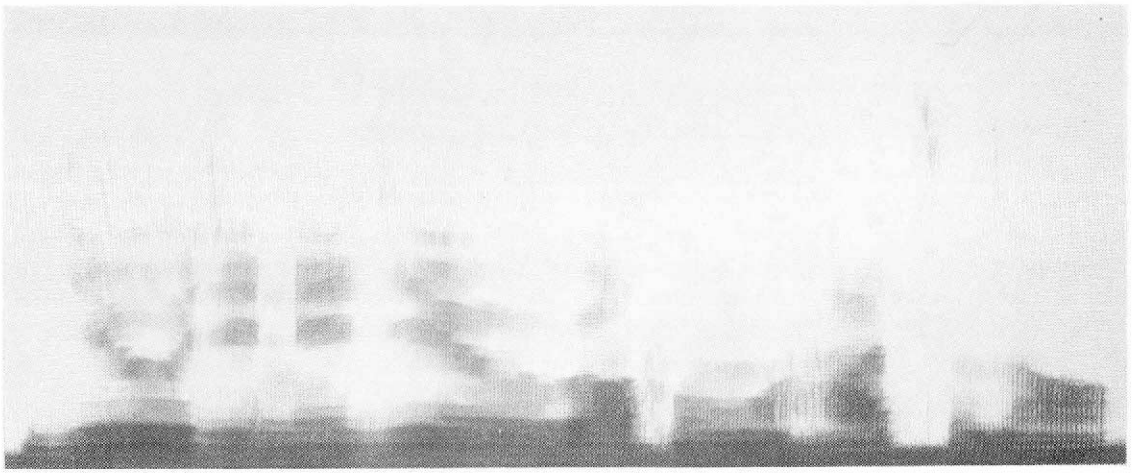
(7)　ありますか



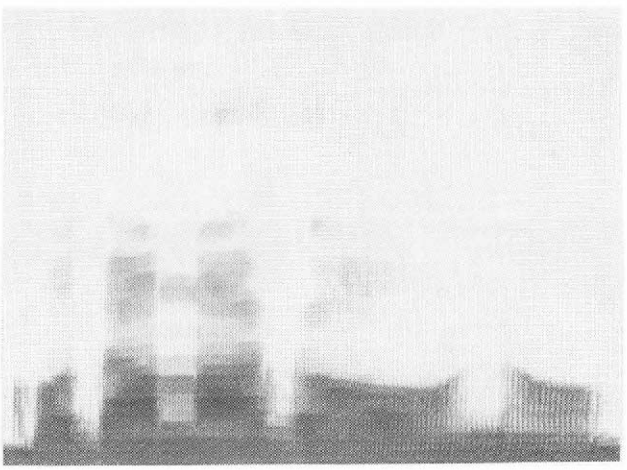
(8)　三島を



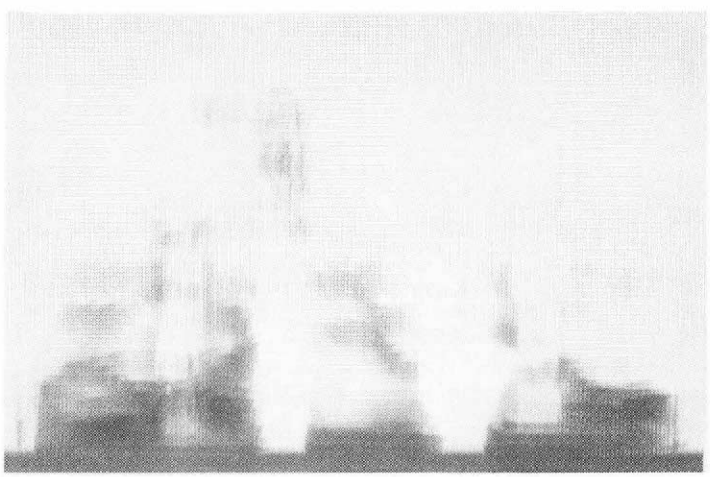
(8)　10時25分発の



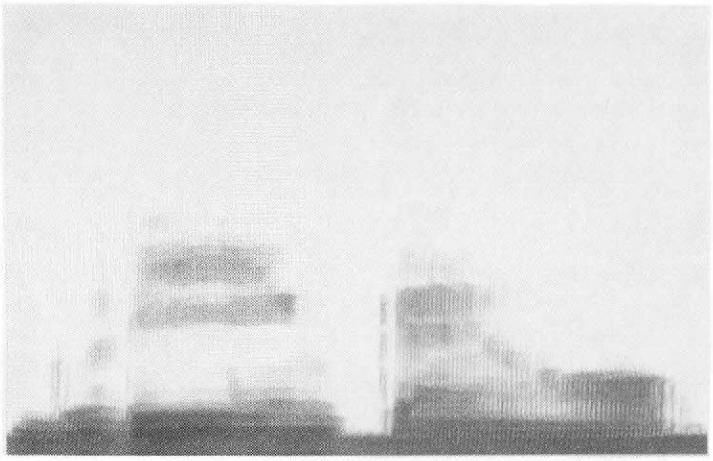
(8)　こだま号の



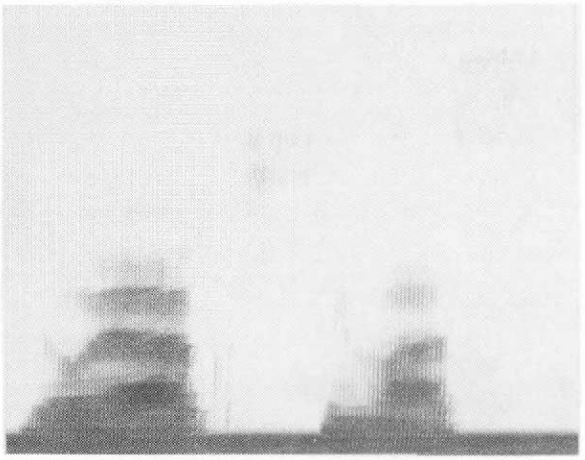
(8)　豊橋ゆきの



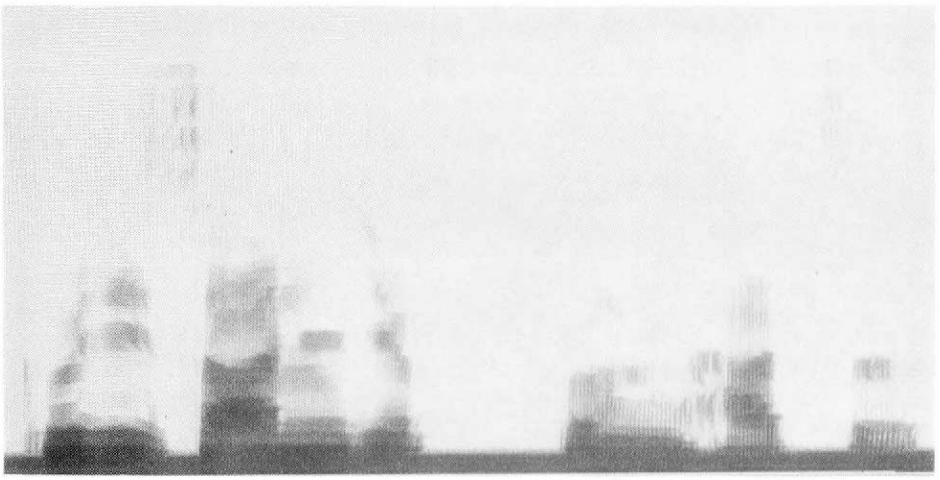
(8)　グリーン券を



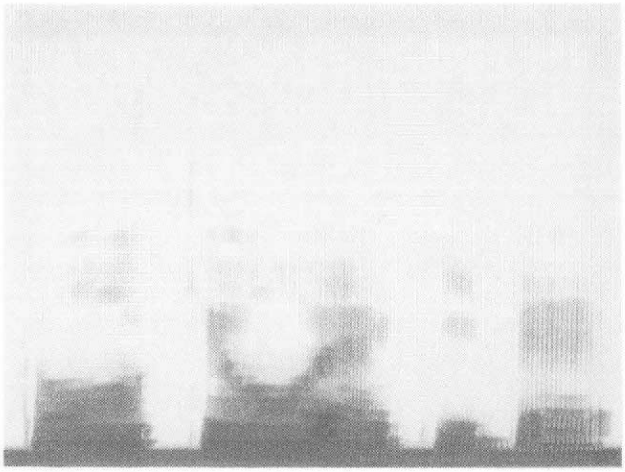
(8)　予約します



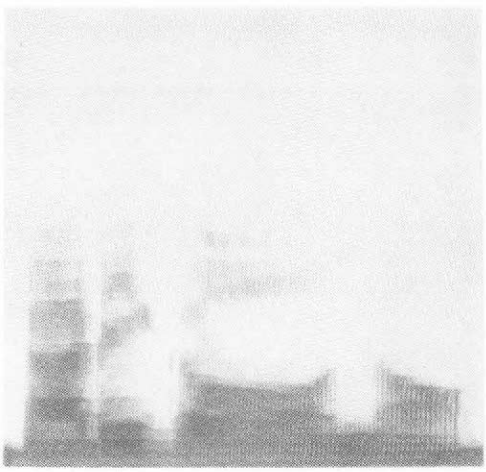
(9)　9時30分発



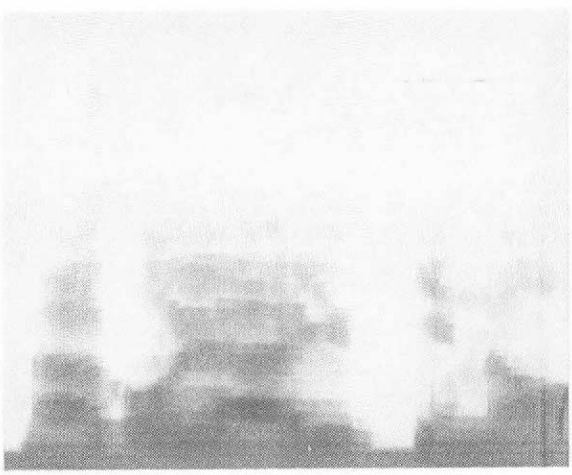
(9)　東京駅で



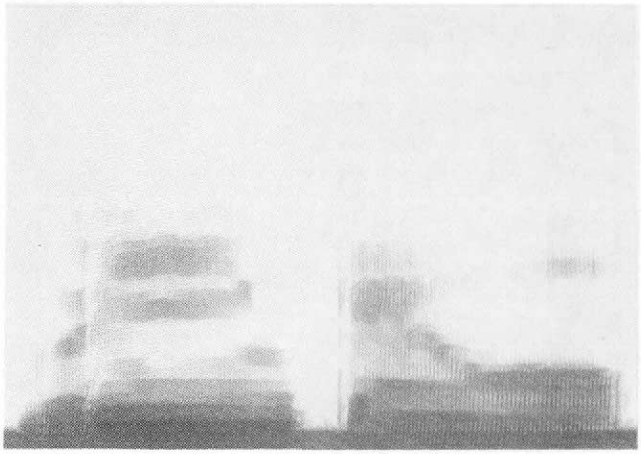
(9)　ひかり号の



(9)　名古屋ゆきの

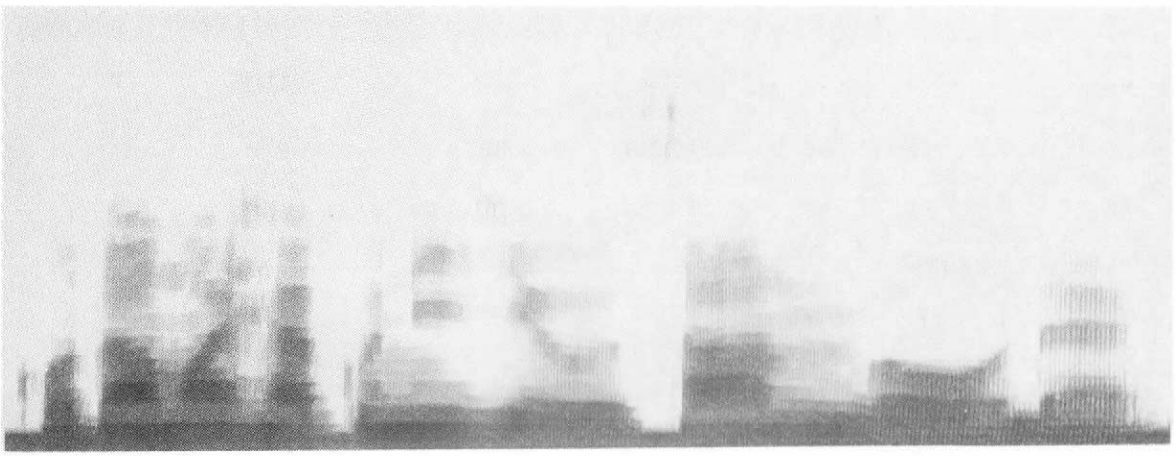


(9)　グリーン券を

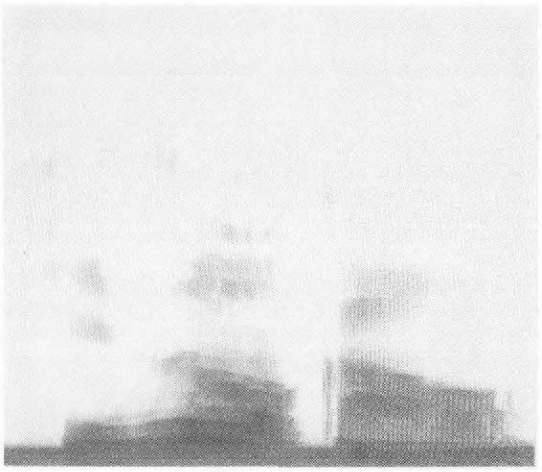


(9)　9枚

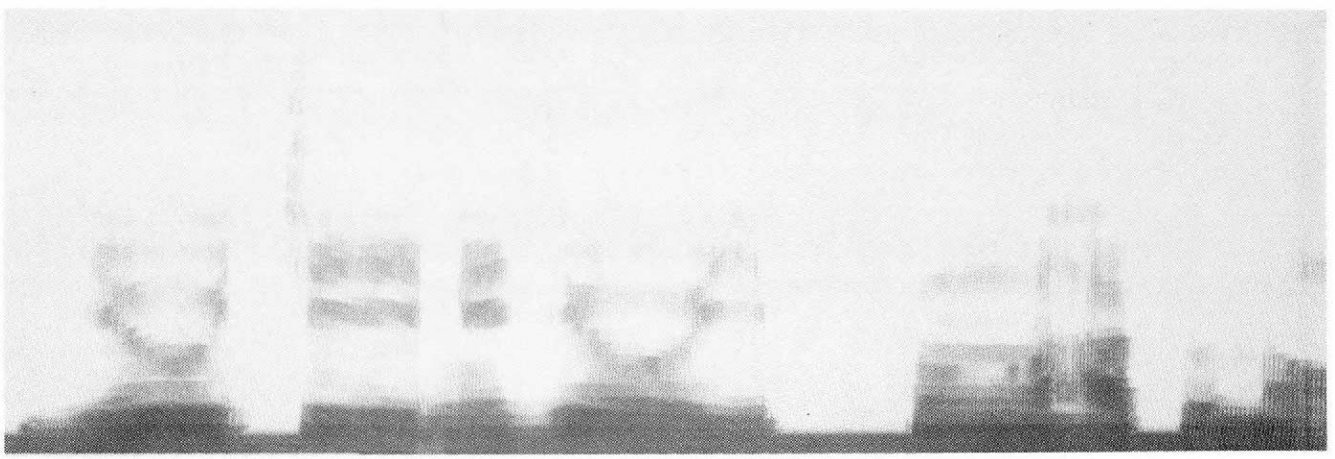
(10)　こだま123号で



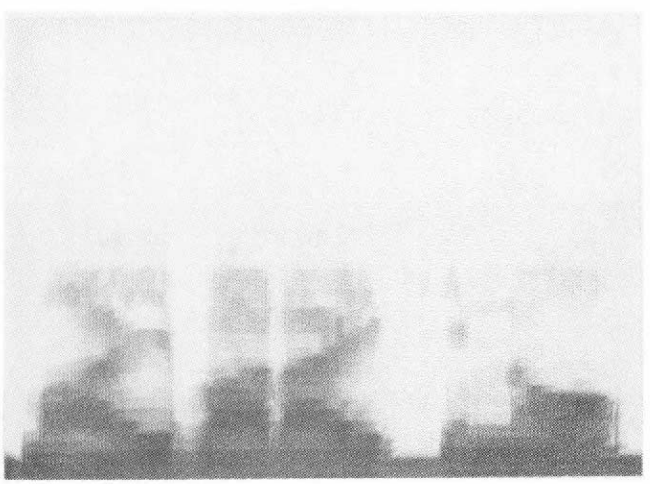
⑽　静岡を



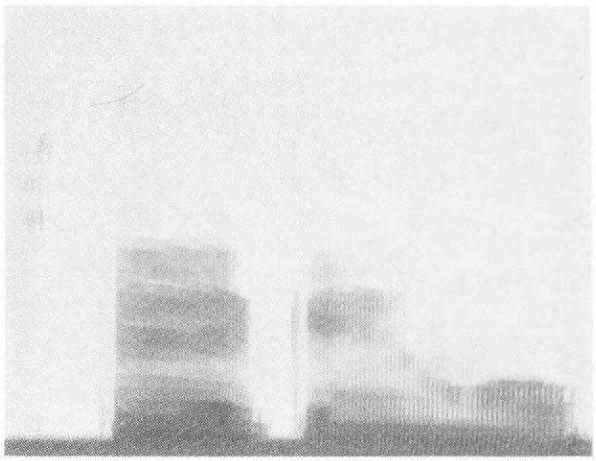
⑽　１１時２１分発の



⑽　米原いきの

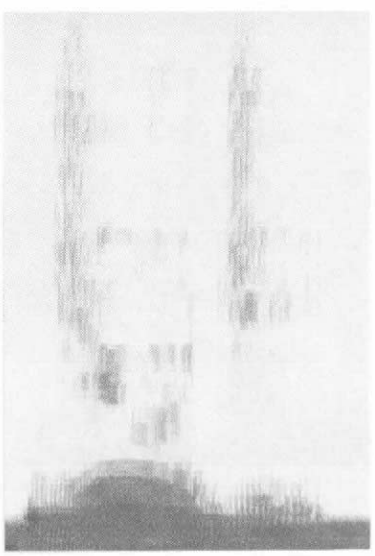


(10)　指定券を

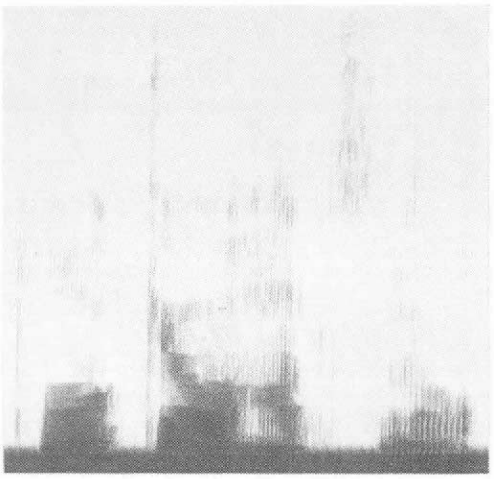


(10)　5枚

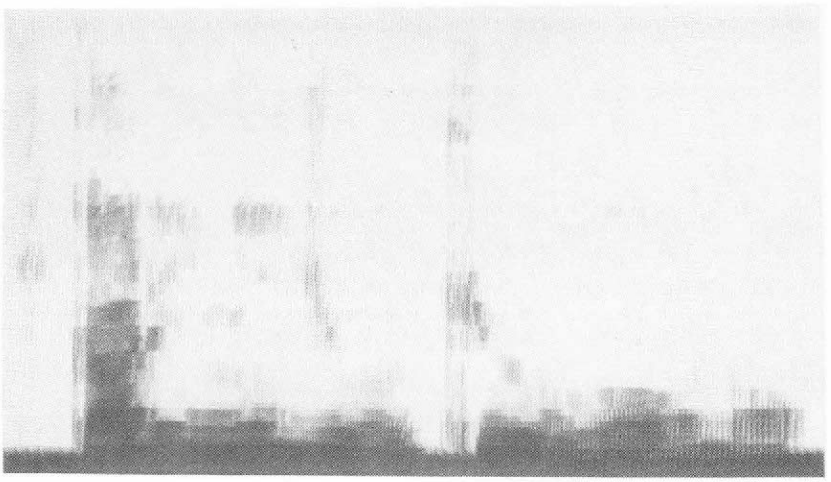
(11)　10時



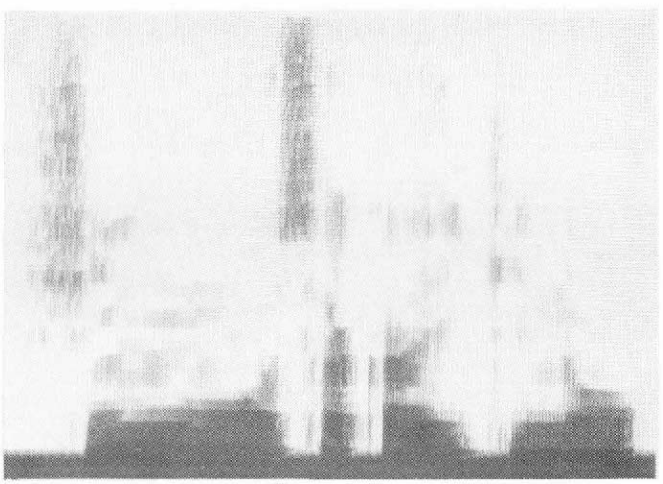
(11)　東京発の



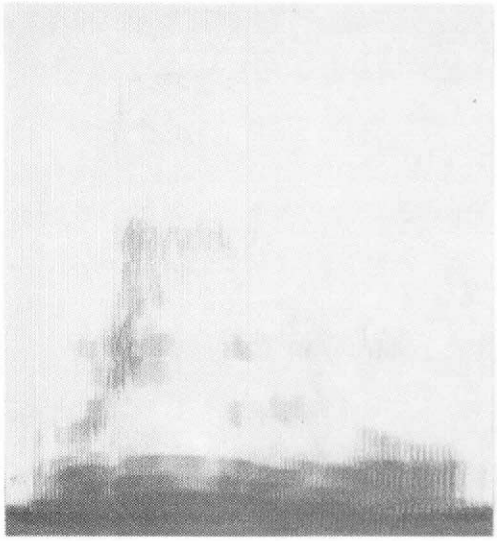
(11)　ひかり２９号の



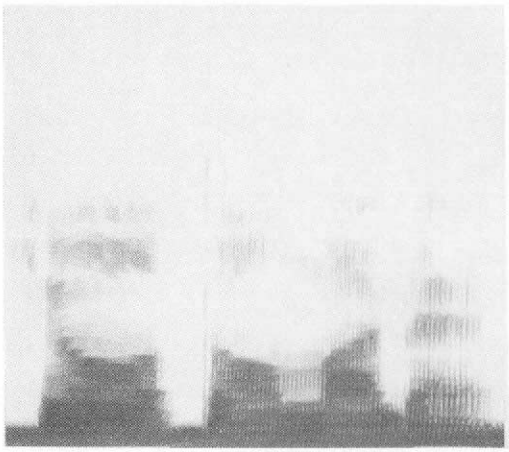
(11)　新大阪いきの



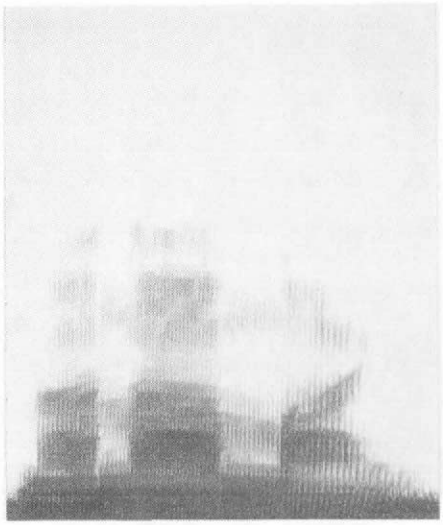
(11)　グリーンを



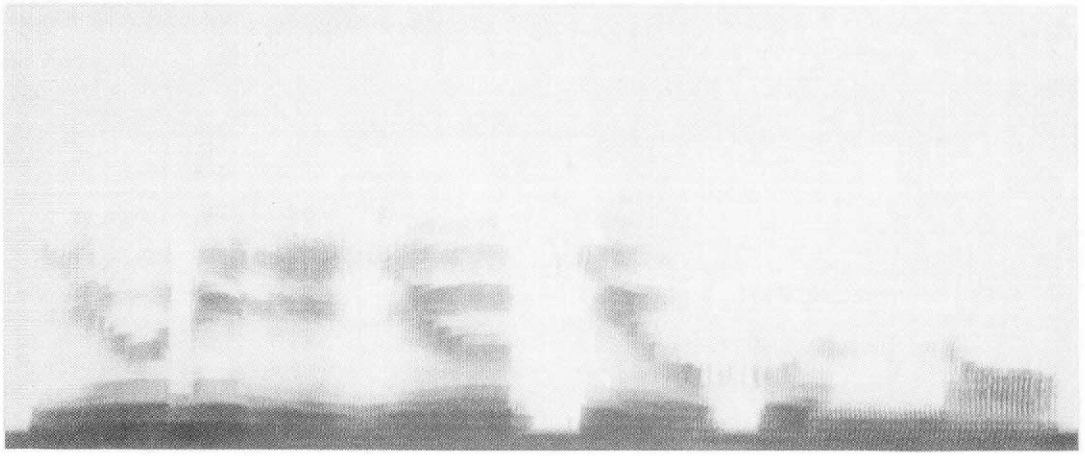
(12)　京都まで



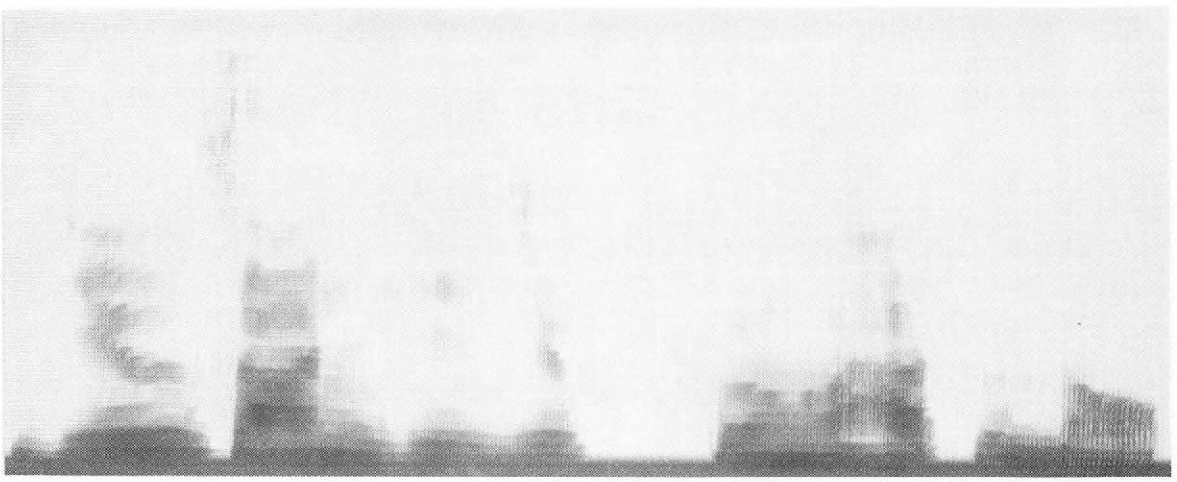
(12)　7 枚



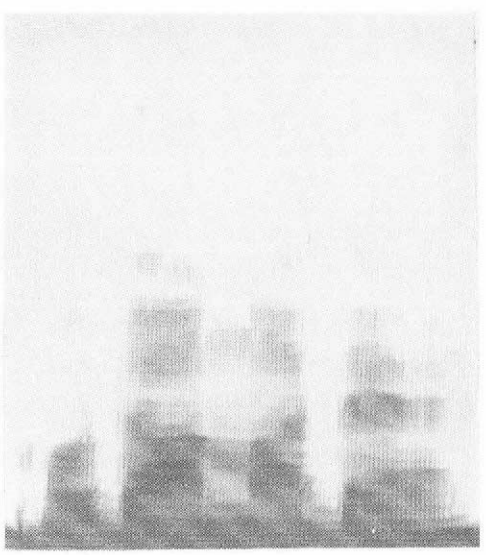
(13)　１２時１９分の



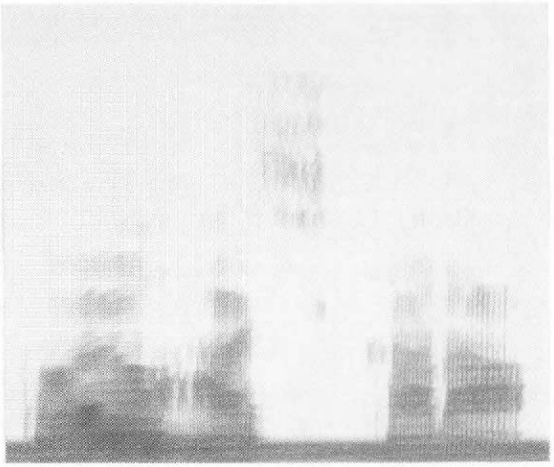
(15)　１３時１０分発の



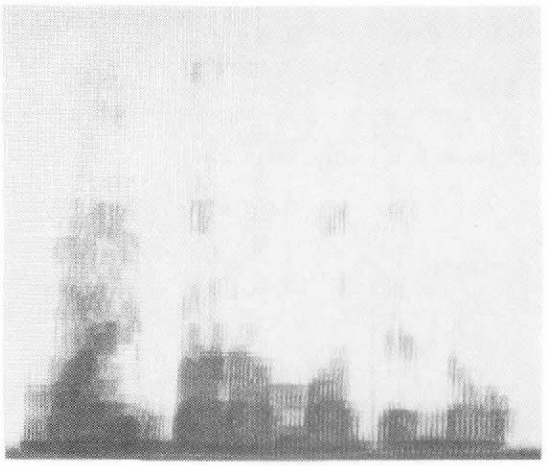
(15)　こだまで



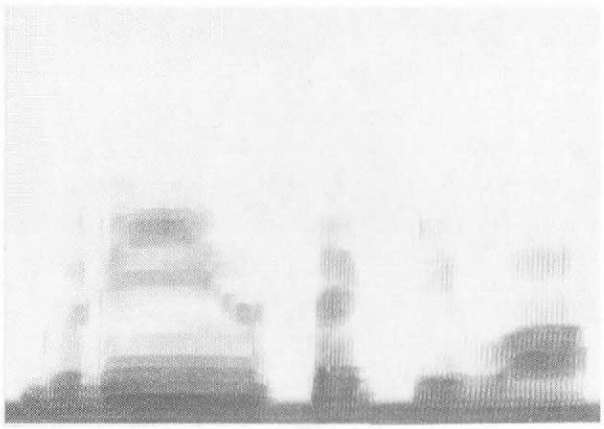
⑮　豊橋から



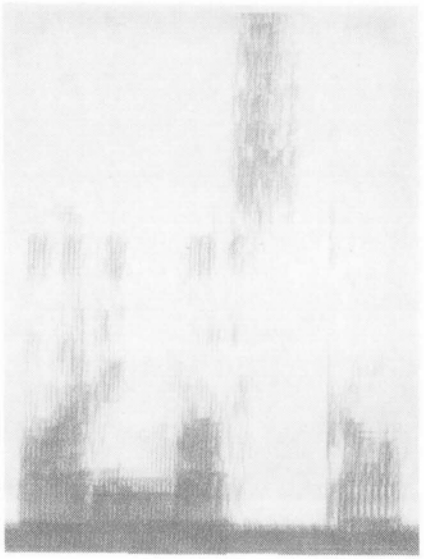
⑮　米原までの



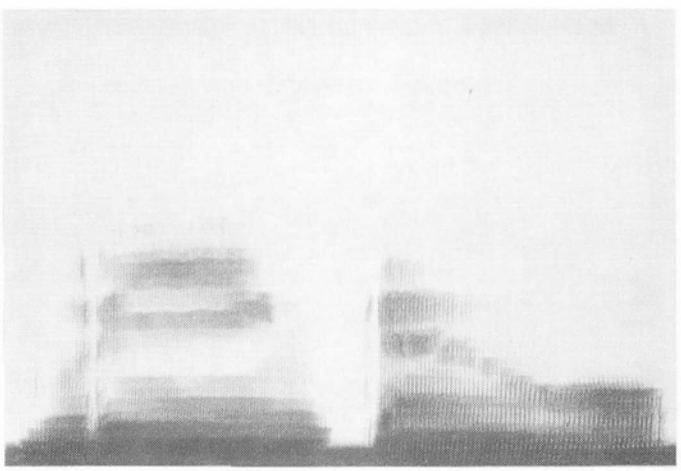
⑮　グリーン席は



(15)　ありますか

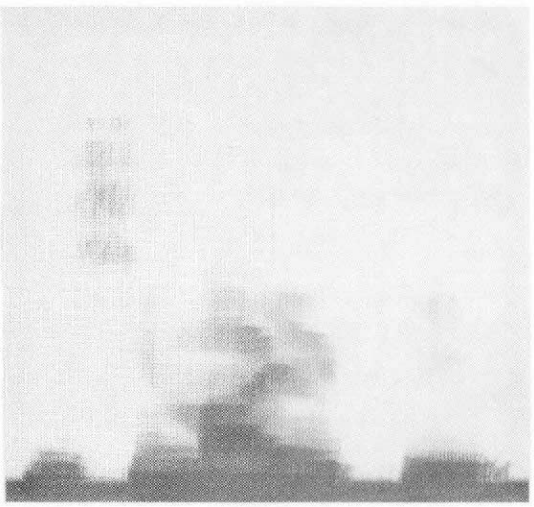


(16)　グリーン券を

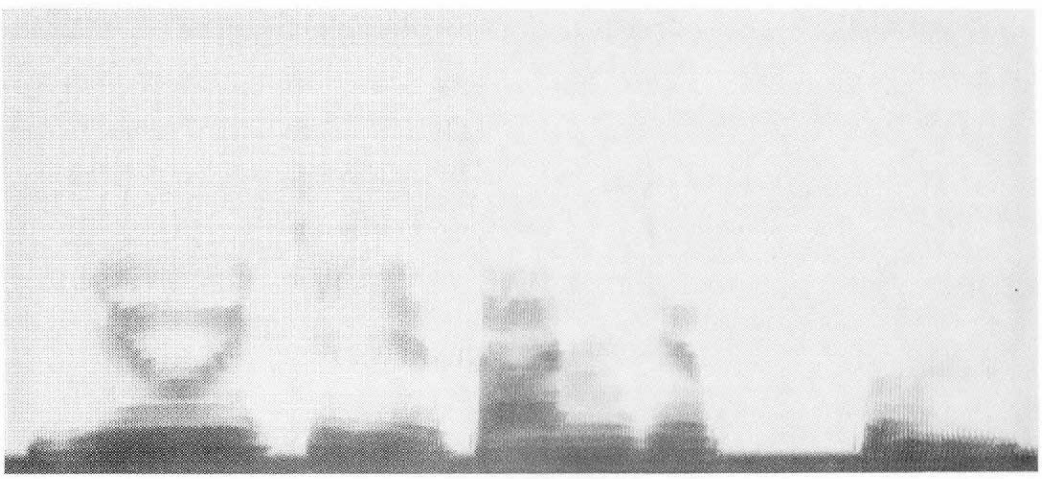


(16)　5枚

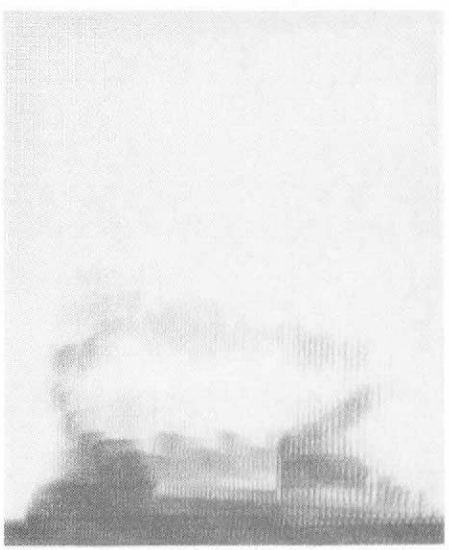
(17)　三島ゆき



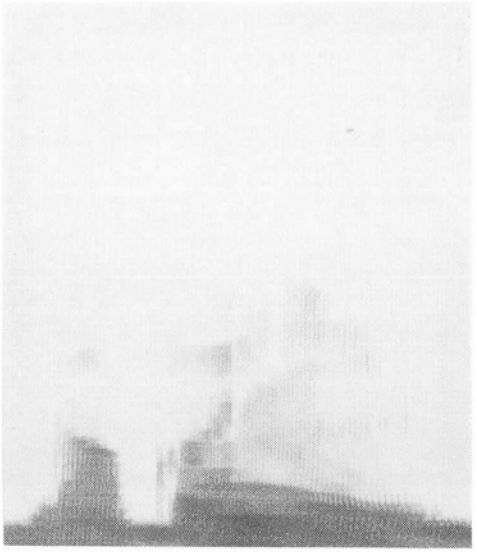
(18)　１１時３０分



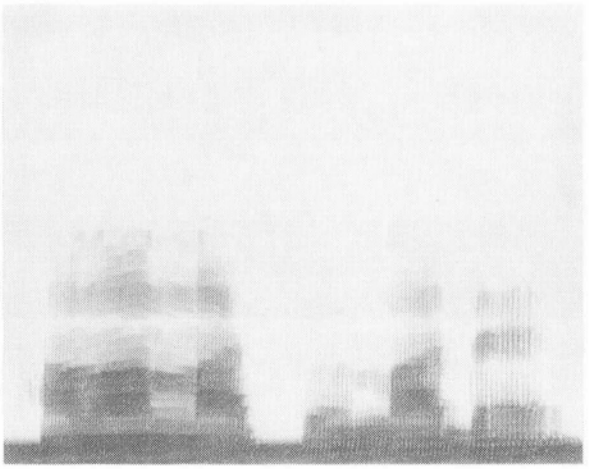
(19)　４枚



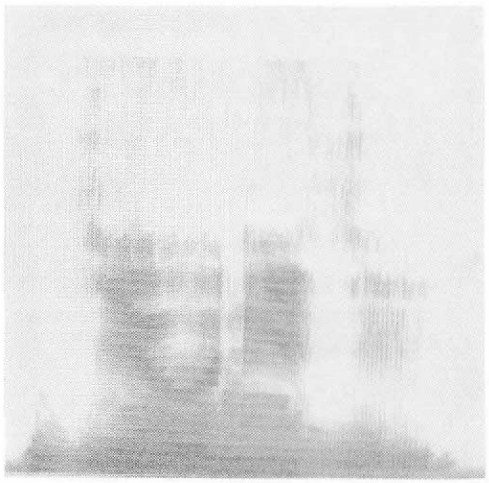
⑳　６人



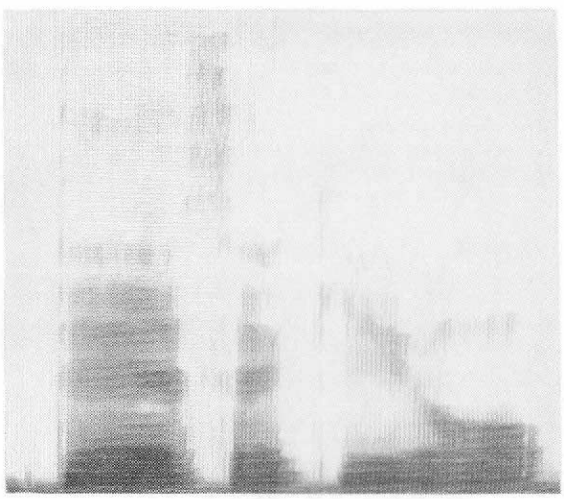
⑳　浜松まで



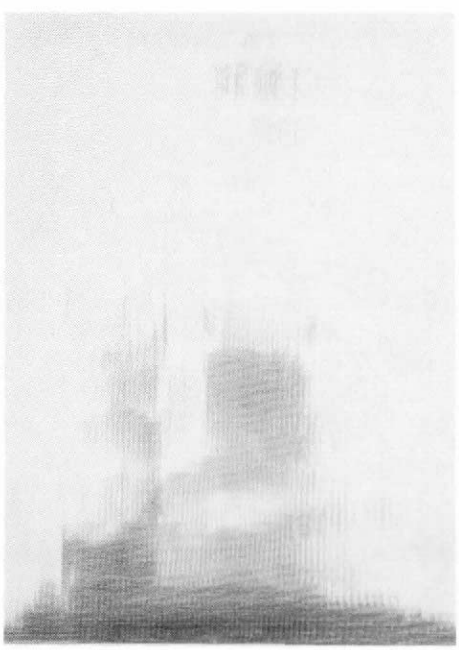
(21)　１２時



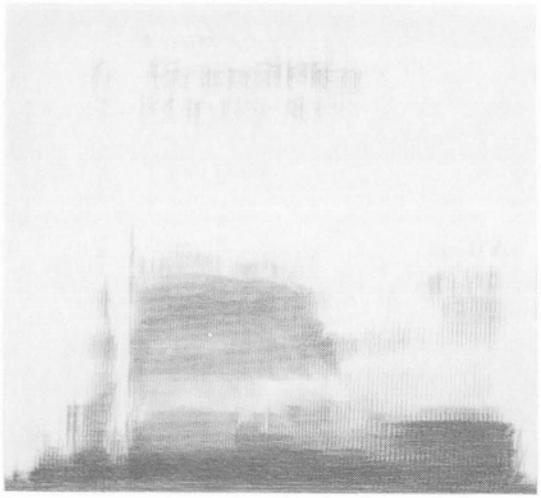
(22)　指定席を



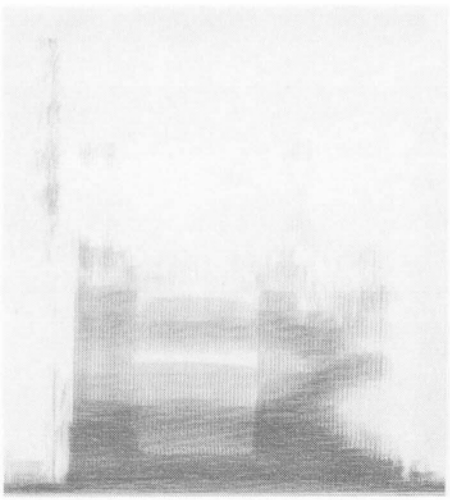
(22)　５人



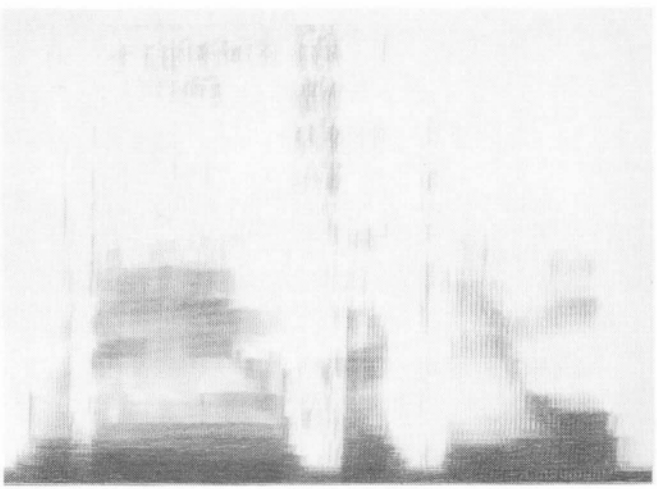
(24)　グリーンを



(24)　3枚

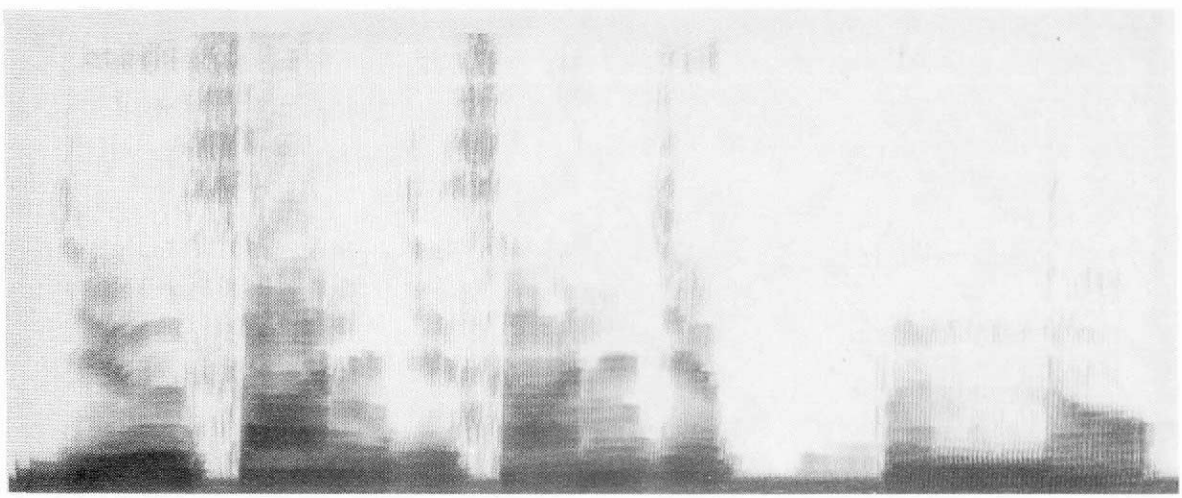


(26)　グリーン席を

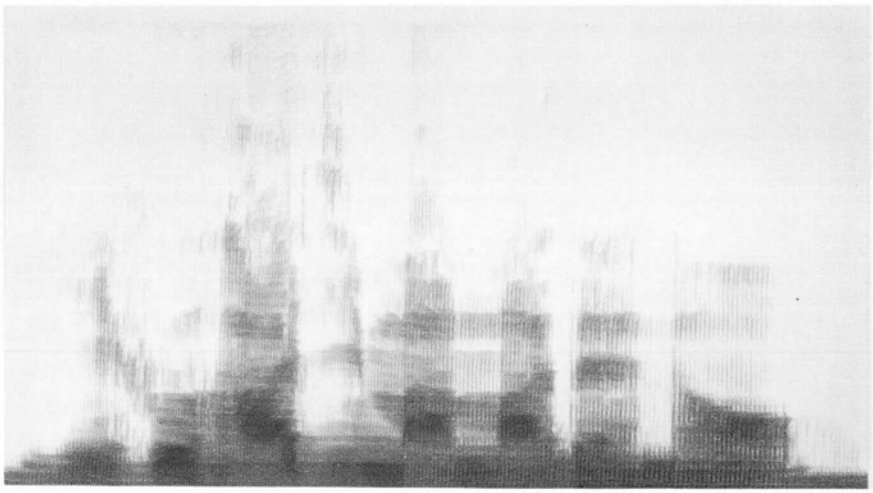


(26)　2枚

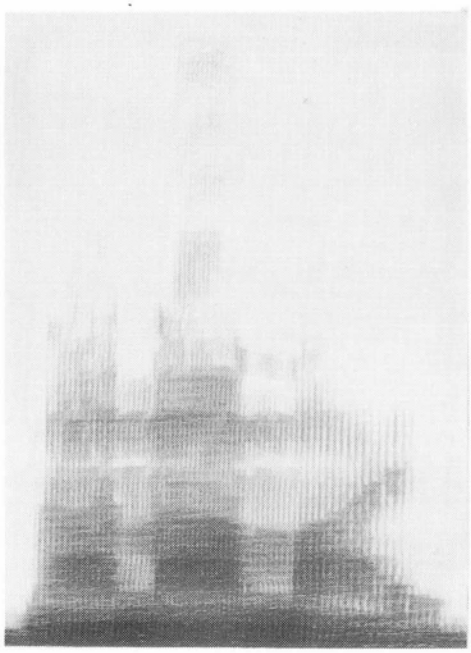
(27)　13時30分の



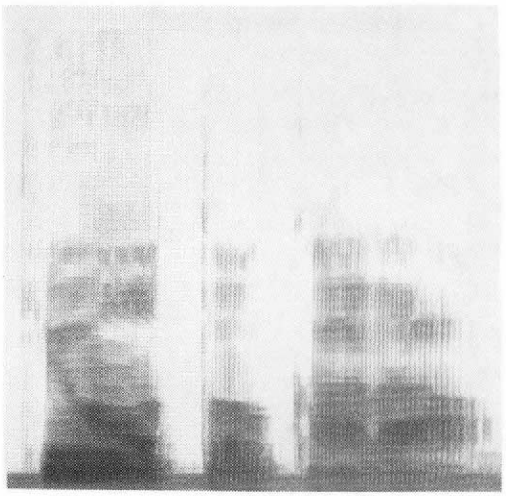
(28)　岐阜羽島までの



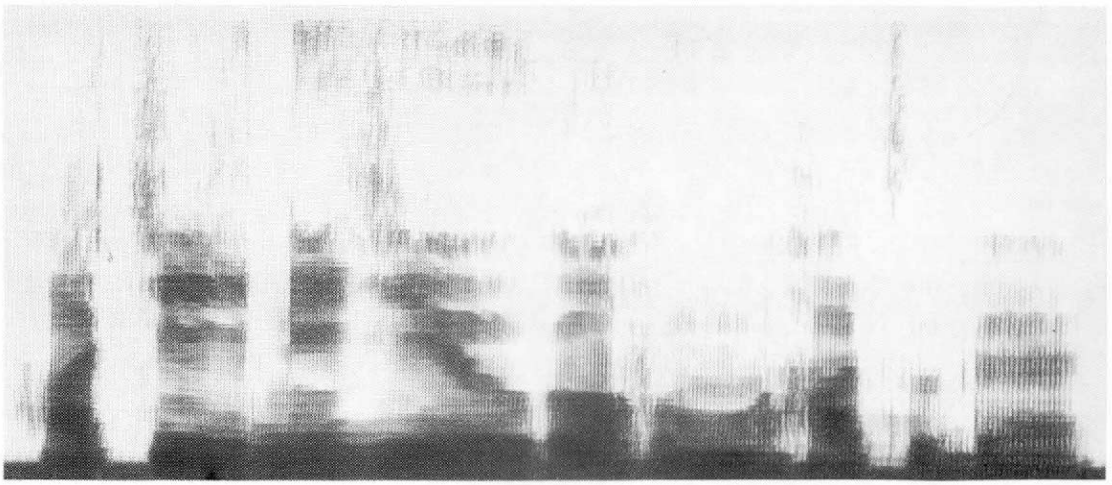
(28)　7枚



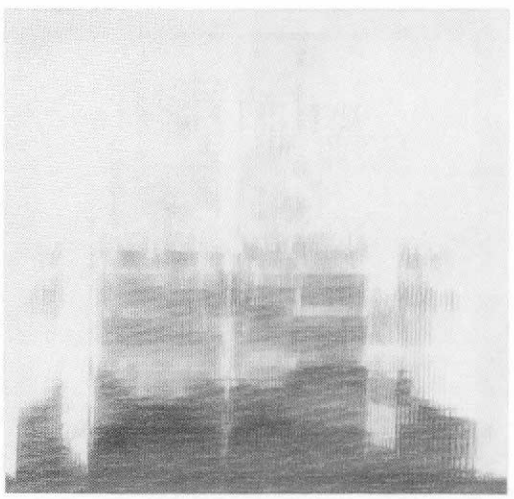
(29)　京都から



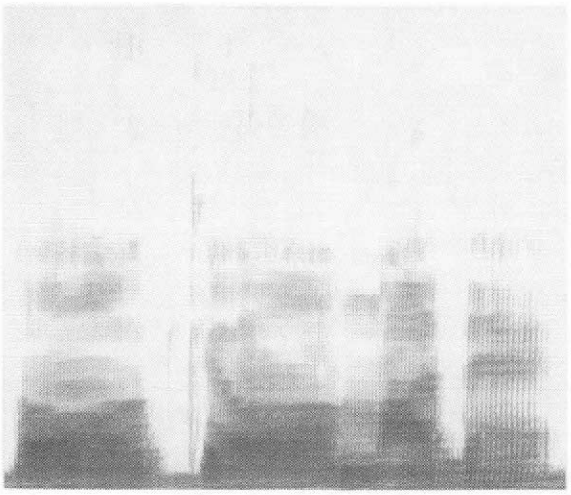
(30)　8時25分発で



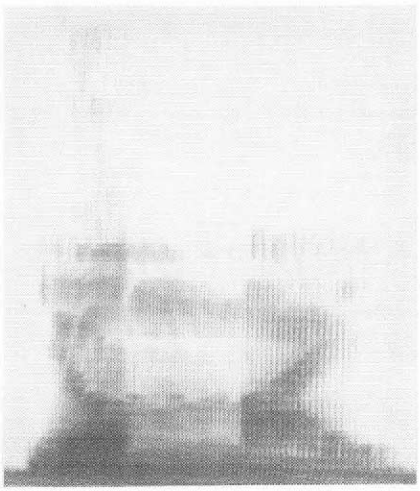
(31)　小田原への



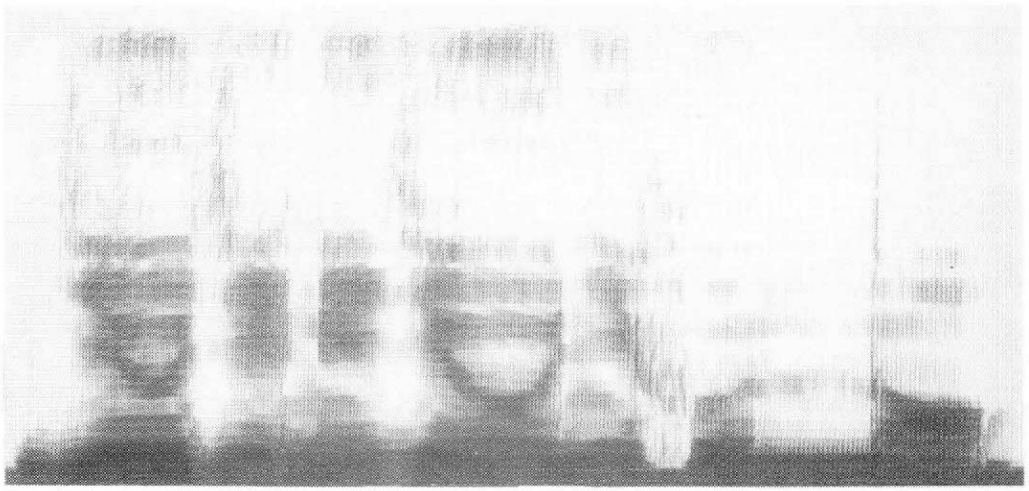
(32)　東京まで



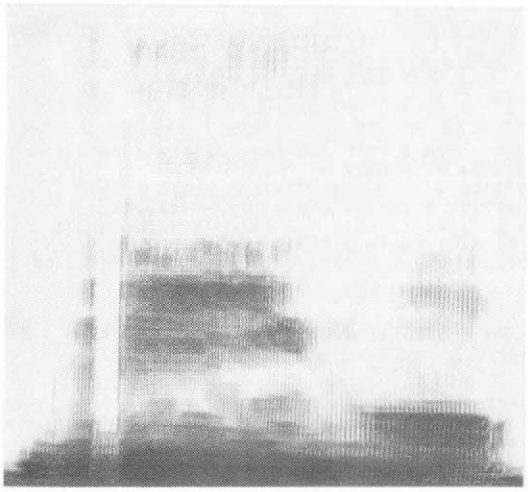
(32)　4枚



(33)　10時22分の

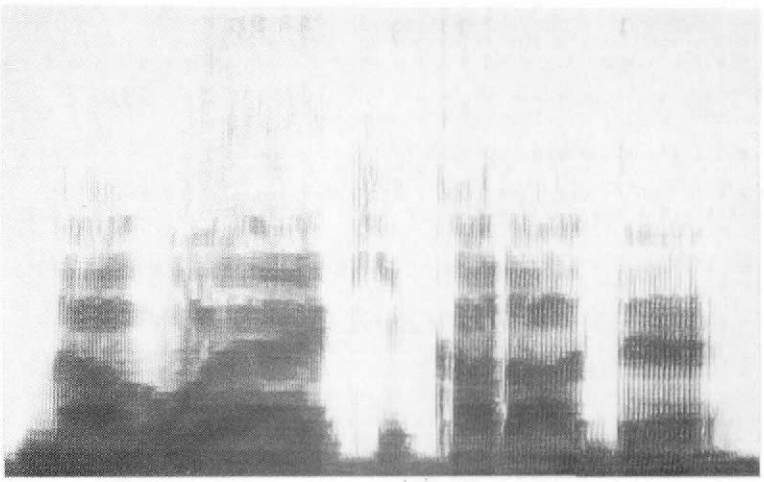


(33)　グリーンを



(33)　4人

(34)　名古屋駅からで



付録 4. 会話音声認識システム（第２次）の言語情報

図Ａ４－１　プラグマティクスのリスト表現

```
PRAGMATICS
( SEM ( YES-NO ) ( * YES-NO ) SEM ( VERB ) SEM ( LATTICE 4 20 ) ( * VERB ) SEM (
 SYNTAX 1 ) SEM ( LATTICE 4 24 )( * DATE )
SEM ( SYNTAX 2 ) SEM ( LATTICE 6 26 ) ( * STARTING-STATION )
 SEM ( SYNTAX 3 ) SEM ( LATTICE 8 26 ) ( * ARRIVING-STATION
 ) SEM ( SYNTAX 4 ) SEM ( LATTICE 12 64 ) ( * STARTING-TIME ) SEM ( SYNTAX 5 )
SEM ( LATTICE 10 64 )( * NAME-OF-TRAIN ) SEM ( SYNTAX 6 ) SEM ( LATTICE 2 18 )(
 * SEAT-CLASS ) SEM ( SYNTAX 7 ) SEM ( LATTICE 4 14 ) ( * NUMBER-OF-TICKETS ) )
```

図Ａ４－２　構文のリスト表現

(Date , Starting Station , Arriving Station , Starting Time , Train Name , Seat Class , Number of Tickets , Verb and Yes-No)

```
DATE
( ( OR ( * TSUITACHI ) ( * FUTSUKA ) ( * MIKKA ) ( * YOKKA ) ( * ITSUKA ) (
 * MUIKA ) ( * NANOKA ) ( * YOKA ) ( * KOKONOKA ) ( * TOKA ) ( * HATSUKA )
( * KYOASU )
( ( OPT ( * 2 ) ) ( * 10 ) ( OR ( ( OR ( * 1 ) ( * 2 ) ( * 3 )( * 5 ) ( * 6 )
 ( * SHICHI ) ( * 7 )
  ( * 8 ) ( * KU ) ) ( * NICHI ) ) ( * YOKKA ))) ( ( * 3 ) ( * 10 ) (
OPT ( * 1 ) ) ( * NICHI ) ) ) ( * AX ) )
STARTING-STATION
( ( * EKIMEI ) ( OPT ( * EKI ) )  ( OR ( * KARA )( * HATSU ) ( * YORI )) (
 * AX ) )
ARRIVING-STATION
( ( * EKIMEI ) ( OPT ( * EKI ) ) ( OR ( * MADE ) ( * YUKI ) ( * IKI )
( SEM ( LATTICE 3 5 ) ( * E ) ) ) ( * AX ) )
STARTING-TIME
( ( * SUJI6-22 ) ( * JI ) ( OPT ( * PAUSE ) ) ( * SUJI0-59FUN ) ( OPT ( * HATSU
 ) ) ( * AX ) )
NAME-OF-TRAIN
 160
( ( OR ( ( * HIKARI ) ( OPT ( * PAUSE ) ) ( * SUJI1-199 ) ) ( ( * KODAMA ) ( OPT
 ( * PAUSE ) ) ( * SUJI200-299 ) ) ) ( * GO ) ( * AX ) )
SEAT-CLASS
( ( OR ( * SHITEI )( * FUTSU )( * GREEN ) ) ( OPT ( OR ( * KEN ) ( * SEKI ) ) )
( * AX ) )
NUMBER-OF-TICKETS
( ( * SUJI1-9 ) ( * MAI ) ( * AX ) )
YES-NO
( OR ( * HAI ) ( * IIE ) ( ( * SO ) ( * DESS ) ) ( ( * CHIGAI ) ( * MASS ) ) )
VERB
( OR ( ( OR ( ( OR ( * YOYAKU ) ( * ONEGAI ) ) ( OPT ( * ITA ) )  ( *
 SHI )) ( ( * MOSHIKOMI )( OPT ( ( * ITA ) ( * SHI )) ) )) ( * MASS )
   )( ( * ARI ) ( OR ( * MASU ) ( * MASEN ) ) ( * KA ) )
( * DESS ) )
EKIMEI
( OR ( * TOKYO ) ( * SHINYOKOHAMA ) ( * ODAWARA ) ( * ATAMI ) ( * MISHIMA )
( * SHIZUOKA ) ( * HAMAMATSU ) ( * TOYOHASHI ) ( * NAGOYA ) ( * GIFUHASHIMA ) (
 * MAIBARA ) ( * KYOTO ) ( * SHINOSAKA ) ( * SHINKOBE ) ( * NISHIAKASHI )( *
HIMEJI ) ( * AIOI ) ( * OKAYAMA ) ( * SHINKURASHIKI ) ( * FUKUYAMA ) ( * MIHARA
 ) ( * HIROSHIMA ) ( * SHINIWAKUNI ) ( * TOKUYAMA ) ( * OGORI ) ( * SHINSHIMONOS
EKI ) ( * KOKURA ) ( * HAKATA ) )
AX
( OPT ( SEM ( LATTICE 1 6 ) OR ( * NO ) ( * O ) ( * DE ) ( * NOO ) ( * NODE ) (
 * WA ) ( * GA ) ( * NOWA ) ) )
SUJI6-22
( OR ( * 6 ) ( * 7 ) ( * 8 )( * KU ) ( ( * 10 ) ( OPT ( OR ( * SUJI1-9 ) ( * YO
 )   ( * KU ) ) ) ) ( ( * 2 ) ( * 10 ) ( OPT ( OR ( * 1 ) ( * 2 ) ) )
 ) )
SUJI1-9
( OR ( * 1 ) ( * 2 ) ( * 3 ) ( * 4 ) ( * 5 ) ( * 6 ) ( * 7 ) ( * 8 ) ( * 9 ) )
SUJI0-59FUN
( OR (( OR ( * 0 ) ( * RE )) ( * FUN )) (( OPT (( OPT ( * SUJI2-5 ) ) ( * 10 ) )
 ) ( OR ( ( OR ( * 1 ) ( * RO ) ( * HA ) ) ( * PPUN ) ) ( ( OR ( * 2 ) ( * 5 ) (
 * 7 ) ( * 8 )
  ( * 9 ) ) ( * FUN ) ) ( ( OR ( * 3 ) ( * 4 ) ) ( * PUN ) ) ) ) ( ( OPT (
 * SUJI2-5 ) ) ( OR ( * ZI ) ( * ZYU ) ) ( * PPUN ) ) )
SUJI2-5
( OR ( * 2 ) ( * 3 ) ( * 4 ) ( * 5 ) )
SUJI1-199
( OR ( * SUJI1-9 ) ( ( OR (( OPT ( OR ( OR ( * 2 ) ( * 9 ) ) ( ( * 100 ) (
 OPT ( OR ( * 2 ) ( * 5 ) ( * 6 ) ( * 7 ) ( * 8 ) ( * 9 ) ) ) ) ))( * 10 ) ) ( *
 100 ) ) ( OPT ( * SUJI1-9 ) ) ) )
SUJI200-299
( ( * 2 ) ( * 100 ) ( OPT ( ( OPT ( OR ( * 2 )( * 3 ) ( * 6 ) ( * 5 ) ( * 8 ) (
* 9 ) ) ) ( * 10 ) ) ) ( OPT ( * SUJI1-9 ) ) )
KYOASU
( OR ( * KYO ) ( * HONJITSU ) ( * ASU ) ( * ASHITA ) ( * ASATTE ) )
MASS
( OR ( * MAS ) ( * MASU ) )
DESS
( OR ( * DES ) ( * DESU ) )
NICHI
( OR ( * NITI ) ( * NNCHI ) )
```

図Ａ４－３　単語辞書のリスト表現

```
TSUITACHI
( T U I T A T I )
FUTSUKA
( H U T U K A )
MIKKA
( M I KK A )
YOKKA
( Y O KK A )
ITSUKA
( I T U K A )
MUIKA
( M U I K A )
NANOKA
( N A N O K A )
YOKA
( Y OO K A )
KOKONOKA
( K O K O N O K A )
TOKA
( T OO K A )
HATSUKA
( H A T U K A )
1
( I T I )
2
( N I )
3
( S A NN )
4
( Y O NN )
5
( G O )
6
( R O K U )
7
( N A N A )
8
( H A T I )
9
( KY UU )
10
( ZY UU )
YO
( Y O )
SHICHI
( H I T I )
KU
( K U )
I
( I )
RO
( R O )
HA
( H A )
ZI
( Z I )
ZYU
( ZY U )
100
( HY A K U )
0
( R E I )
RE
( R EE )
TOKYO
( T OO KY OO )
SHINYOKOHAMA
( S I NN Y O K O H A M A )
ODAWARA
( O D A W A R A )
ATAMI
( A T A M I )
MISHIMA
( M I S I M A )
SHIZUOKA
( S I Z U O K A )
HAMAMATSU
( H A M A M A T U )
TOYOHASHI
( T O Y O H A S I )
NAGOYA
( N A G O Y A )
GIFUHASHIMA
( G I H U H A S I M A )
MAIBARA
( M A I B A R A )
KYOTO
( KY OO T O )
SHINOSAKA
( S I NN OO S A K A )
SHINKOBE
( S I NN K OO B E )
NISHIAKASHI
( N I S I A K A S I )
HIMEJI
( H I M E Z I )
AIOI
( A I O I )
OKAYAMA
( O K A Y A M A )
SHINKURASHIKI
( S I NN K U R A S I K I )
FUKUYAMA
( H U K U Y A M A )
MIHARA
( M I H A R A )
HIROSHIMA
( H I R O S I M A )
SHINIWAKUNI
( S I NN I W A K U N I )
TOKUYAMA
( T O K U Y A M A )
OGORI
( O G OO R I )
SHINSHIMONOSEKI
( S I NN S I M O N O S E K I )
KOKURA
( K O K U R A )
HAKATA
( H A K A T A )
EKI
( E K I )
NITI
( N I T I )
KARA
( K A R A )
HATSU
( H A T U )
YORI
( Y O R I )
NO
( N O )
O
( O )
DE
( D E )
NOO
( N O O )
NODE
( N O D E )
WA
( W A )
GA
( G A )
MADE
( M A D E )
YUKI
( Y U K I )
IKI
( I K I )
E
( E )
JI
( Z I )
FUN
( H U NN )
PUN
( P U NN )
PPUN
( PP U NN )
HIKARI
( H I K A R I )
KODAMA
( K O D A M A )
GO
( G OO )
GREEN
( G U R II NN )
FUTSU
( H U T UU )
SHITEI
( S I T E I )
KEN
( K E NN )
SEKI
( S E K I )
MAI
( M A I )
HAI
( H A I )
IIE
( I I E )
YOYAKU
( Y O Y A K U )
ONEGAI
( O N E G A I )
ITA
( I T A )
SHI
( S I )
MASU
( M A S U )
ARI
( A R I )
MASEN
( M A S E NN )
KA
( K A )
DESU
( D E S U )
DES
( D E )
MAS
( M A )
PAUSE
( ** )
MOSHIKOMI
( M OO S I K O M I )
NOWA
( N O W A )
KYO
( KY OO )
HONJITSU
( H O NN Z I T U )
ASU
( A S U )
ASHITA
( A S I T A )
ASATTE
( A S A KK E )
NNCHI
( NN T I )
SO
( S OO )
CHIGAI
( T I G A I )
```

図Ａ４－４　時　刻　表

****

SHINKANSEN JIKOKUHYO　　KUDARI　PAGE　1

****

	HIKARI 19 GO	KODAMA 205 GO	KODAMA 213 GO	HIKARI 91 GO	KODAMA 217 GO	HIKARI 17 GO	KODAMA 221 GO	HIKARI 21 GO	KODAMA 201 GO	HIKARI 101 GO	HIKARI 151 GO	KODAMA 203 GO	HIKARI 121 GO	KODAMA 207 GO	HIKARI 1 GO
TOKYO								6:00	6:04	6:12	6:24	6:28	6:48	6:52	7:00
SHINYOKOHAMA								-	6:22	-	-	6:46	-	7:10	-
ODAWARA								-	6:46	-	-	7:10	-	7:34	-
ATAMI								-	6:57	-	-	7:21	-	7:45	-
MISHIMA								-	7:11	-	-	7:35	-	7:59	-
SHIZUOKA								-	7:39	-	-	8:04	-	8:28	-
HAMAMATSU								-	8:07	-	-	8:31	-	8:55	-
TOYOHASHI								-	8:27	-	-	8:51	-	9:15	-
NAGOYA						7:03		8:03	8:56	8:15	8:27	9:20	8:51	9:44	9:03
GIFUHASHIMA						-		-	9:15	-	-	9:39	-	10:03	-
MAIBARA						-		-	9:33	-	-	9:57	-	10:21	-
KYOTO						7:53		8:53	10:00	9:05	9:17	10:24	9:41	10:48	9:53
SHINOSAKA	6:02	6:06	6:49	7:02	7:41	8:12	8:36	9:12	10:18	9:24	9:37	10:42	10:00	11:06	10:12
SHINKOBE	-	6:22	7:05	-	7:57	-	8:52	-		9:40	9:53		-		-
NISHIAKASHI	-	6:35	7:17	-	8:09	-	9:04	-		-	10:05		-		-
HIMEJI	-	6:50	7:35	-	8:24	-	9:19	-		10:04	10:20		-		-
AIOI	-	7:02	7:47	-	8:35	-	9:30	-		-	10:31		-		-
OKAYAMA	7:02	7:27	8:12	8:02	8:59	9:12	9:54	10:12		10:35	10:53		11:00		11:12
SHINKURASHIKI	-	7:41	8:26	-	9:13	-	10:09	-		10:49			11:15		-
FUKUYAMA	-	7:56	8:41	-	9:28	-	10:26	-		11:04			11:36		-
MIHARA	-	8:11	8:56	-	9:46	-	10:46	-		11:19			11:50		-
HIROSHIMA	8:00	8:39	9:24	9:00	10:14	10:10	11:14	11:10		11:46			12:19		12:10
SHINIWAKUNI	-	8:58	9:43	-	10:33	-	11:33	-		12:05			12:38		-
TOKUYAMA	8:33	9:16	10:03	9:33	10:53	-	11:53	-		12:24			12:58		-
OGORI	-	9:38	10:23	-	11:13	-	12:13	11:58		12:44			13:18		-
SHINSHIMONOSEKI	9:11	10:01	10:46	10:11	11:36	-	12:37	-		13:07			13:42		-
KOKURA	9:24	10:14	10:57	10:24	11:49	11:24	12:51	12:29		13:19			13:56		13:24
HAKATA	9:55	10:46	11:31	10:56	12:21	11:56	13:25	13:01		13:51			14:30		13:56

	KODAMA 209 GO	HIKARI 153 GO	HIKARI 105 GO	KODAMA 211 GO	HIKARI 191 GO	HIKARI 3 GO	KODAMA 215 GO	HIKARI 155 GO	KODAMA 219 GO	HIKARI 107 GO	HIKARI 23 GO	KODAMA 223 GO	HIKARI 157 GO	KODAMA 225 GO	HIKARI 109 GO
TOKYO	7:16	7:24	7:36	7:40	7:48	8:00	8:16	8:24	8:40	8:48	9:00	9:16	9:24	9:40	9:48
SHINYOKOHAMA	7:34	-	-	7:58	-	-	8:34	-	8:58	-	-	9:34	-	9:58	-
ODAWARA	7:58	-	-	8:22	-	-	8:58	-	9:22	-	-	9:58	-	10:22	-
ATAMI	8:09	-	-	8:33	-	-	9:09	-	9:33	-	-	10:09	-	10:33	-
MISHIMA	8:23	-	-	8:47	-	-	9:23	-	9:47	-	-	10:23	-	10:47	-
SHIZUOKA	8:52	-	-	9:16	-	-	9:52	-	10:16	-	-	10:52	-	11:16	-
HAMAMATSU	9:19	-	-	9:43	-	-	10:19	-	10:43	-	-	11:19	-	11:43	-
TOYOHASHI	9:39	-	-	10:03	-	-	10:59	-	11:03	-	-	11:39	-	12:03	-
NAGOYA	10:08	9:27	9:39	10:32	9:51	10:03	11:08	10:27	11:32	10:51	11:03	12:08	11:27	12:32	11:51
GIFUHASHIMA	10:27	-	-	10:51	-	-	11:27	-	11:51	-	-	12:27	-	12:51	-
MAIBARA	10:45	-	-	11:09	10:17	-	11:45	-	12:09	-	-	12:45	-	13:09	-
KYOTO	11:12	10:17	10:29	11:36	10:44	10:53	12:12	11:17	12:36	11:41	11:53	13:12	12:17	13:36	12:41
SHINOSAKA	11:30	10:37	10:48	11:54	11:02	11:12	12:30	11:37	12:54	12:00	12:12	13:30	12:37	13:54	13:00
SHINKOBE		10:53	11:04			-		11:53		12:17	-		12:53		13:17
NISHIAKASHI		11:05	-			-		12:05		-	-		13:05		-
HIMEJI		11:20	11:28			-		12:20		12:45	-		13:20		13:45
AIOI		11:31	-			-		12:31		-	-		13:31		-
OKAYAMA		11:53	11:59			12:12		12:53		13:17	13:12		13:53		14:17
SHINKURASHIKI			12:13			-				13:31	-				14:31
FUKUYAMA			12:28			-				13:46	-				14:46
MIHARA			12:46			-				14:01	-				15:01
HIROSHIMA			13:16			13:10				14:29	14:10				15:29
SHINIWAKUNI			13:35			-				14:48	-				15:48
TOKUYAMA			13:55			-				15:08	-				16:08
OGORI			14:15			-				15:28	14:58				16:28
SHINSHIMONOSEKI			14:38			-				15:51	-				16:51
KOKURA			14:51			14:24				16:04	15:29				17:04
HAKATA			15:25			14:56				16:36	16:01				17:36

	HIKARI 5 GO	HIKARI 127 GO	KODAMA 229 GO	HIKARI 159 GO	KODAMA 233 GO	HIKARI 111 GO	HIKARI 7 GO	KODAMA 237 GO	HIKARI 161 GO	KODAMA 239 GO	HIKARI 113 GO	HIKARI 9 GO	KODAMA 231 GO	HIKARI 163 GO	KODAMA 257 GO
TOKYO	10:00	10:12	10:16	10:24	10:40	10:48	11:00	11:16	11:24	11:40	11:48	12:00	12:16	12:24	12:40
SHINYOKOHAMA	-	-	10:34	-	10:58	-	-	11:34	-	11:58	-	-	12:34	-	12:58
ODAWARA	-	-	10:58	-	11:22	-	-	11:58	-	12:22	-	-	12:58	-	13:22
ATAMI	-	-	11:00	-	11:33	-	-	12:09	-	12:33	-	-	13:09	-	13:33
MISHIMA	-	-	11:23	-	11:47	-	-	12:23	-	12:47	-	-	13:23	-	13:47
SHIZUOKA	-	-	11:52	-	12:16	-	-	12:52	-	13:16	-	-	13:52	-	14:16
HAMAMATSU	-	-	12:19	-	12:43	-	-	13:19	-	13:43	-	-	14:19	-	14:43
TOYOHASHI	-	-	12:39	-	13:03	-	-	13:39	-	14:03	-	-	14:39	-	15:03
NAGOYA	12:03	12:15	13:08	12:27	13:32	12:51	13:03	14:08	13:27	14:32	13:51	14:03	15:08	14:27	15:32
GIFUHASHIMA	-	-	13:27	-	13:51	-	-	14:27	-	14:51	-	-	15:27	-	15:51
MAIBARA	-	-	13:45	-	14:09	-	-	14:45	-	15:09	-	-	15:45	-	16:09
KYOTO	12:53	13:05	14:12	13:17	14:36	13:41	13:53	15:12	14:17	15:36	14:41	14:53	16:12	15:17	16:36
SHINOSAKA	13:12	13:24	14:30	13:37	14:54	14:00	14:12	15:30	14:37	15:54	15:00	15:12	16:30	15:37	16:54
SHINKOBE	-	13:40		13:53		14:17	-		14:53		15:17	-		15:53	
NISHIAKASHI	-	-		14:05		-	-		15:05		-	-		16:05	
HIMEJI	-	14:04		14:20		14:45	-		15:20		15:45	-		16:20	
AIOI	-	-		14:31		-	-		15:31		-	-		16:31	
OKAYAMA	14:12	14:33		14:53		15:17	15:12		15:53		16:17	16:12		16:53	
SHINKURASHIKI	-					15:31	-				16:31	-			
FUKUYAMA	-					15:46	-				16:46	-			
MIHARA	-					16:01	-				17:01	-			
HIROSHIMA	15:10					16:29	16:10				17:29	17:10			
SHINIWAKUNI	-					16:48	-				17:48	-			
TOKUYAMA	-					17:08	-				18:08	-			
OGORI	-					17:28	-				18:29	-			
SHINSHIMONOSEKI	-					17:51	-				18:51	-			
KOKURA	16:24					18:04	17:24				19:04	18:24			
HAKATA	16:56					18:36	17:56				19:36	18:56			

	HIKARI 115 GO	HIKARI 11 GO	KODAMA 251 GO	HIKARI 165 GO	KODAMA 253 GO	HIKARI 123 GO	HIKARI 25 GO	HIKARI 181 GO	KODAMA 259 GO	HIKARI 167 GO	KODAMA 263 GO	HIKARI 117 GO	HIKARI 13 GO	HIKARI 97 GO	KODAMA 267 GO
TOKYO	12:48	13:00	13:16	13:24	13:40	13:48	14:00	14:12	14:16	14:24	14:40	14:48	15:00	15:12	15:16
SHINYOKOHAMA	–	–	13:34	–	13:58	–	–	–	14:34	–	14:58	–	–	–	15:34
ODAWARA	–	–	13:50	–	14:22	–	–	–	14:58	–	15:22	–	–	–	15:58
ATAMI	–	–	14:09	–	14:33	–	–	–	15:09	–	15:33	–	–	–	16:09
MISHIMA	–	–	14:23	–	14:47	–	–	–	15:23	–	15:47	–	–	–	16:23
SHIZUOKA	–	–	14:52	–	15:16	–	–	–	15:52	–	16:16	–	–	–	16:52
HAMAMATSU	–	–	15:19	–	15:43	–	–	–	16:19	–	16:43	–	–	–	17:19
TOYOHASHI	–	–	15:39	–	16:03	–	–	–	16:39	–	17:03	–	–	–	17:39
NAGOYA	14:51	15:03	16:08	15:27	16:32	15:51	16:03	16:15	17:08	16:27	17:32	16:51	17:03	17:15	18:08
GIFUHASHIMA	–	–	16:27	–	16:51	–	–	–	17:27	–	17:51	–	–	–	18:27
MAIBARA	–	–	16:45	–	17:09	–	–	–	17:45	–	18:09	–	–	–	18:45
KYOTO	15:41	15:53	17:12	16:17	17:36	16:41	16:53	17:05	18:12	17:17	18:36	17:41	17:53	18:05	19:12
SHINOSAKA	16:00	16:12	17:30	16:37	17:54	17:00	17:12	17:24	18:30	17:37	18:54	18:00	18:12	18:22	19:30
SHINKOBE	16:17	–		16:53		–	–	17:40		17:53		18:17	–		
NISHIAKASHI	–	–		17:05		–	–	–		18:05		–	–		
HIMEJI	16:45	–		17:20		–	–	18:04		18:20		18:45	–		
AIOI	–	–		17:31		–	–	–		18:31		–	–		
OKAYAMA	17:17	17:12		17:53		18:00	18:12	18:33		18:53		19:17	19:12		
SHINKURASHIKI	17:31	–				18:15	–					19:31	–		
FUKUYAMA	17:46	–				18:35	–					19:46	–		
MIHARA	18:01	–				18:50	–					20:01	–		
HIROSHIMA	18:29	18:10				19:19	19:10					20:29	20:10		
SHINIWAKUNI	18:48	–				19:38	–					20:48	–		
TOKUYAMA	19:08	–				19:58	–					21:08	–		
OGORI	19:28	–				20:18	19:58					21:28	–		
SHINSHIMONOSEKI	19:51	–				20:42	–					21:51	–		
KOKURA	20:04	19:24				20:56	20:29					22:04	21:24		
HAKATA	20:26	19:56				21:30	21:01					22:36	21:56		

	HIKARI 169 GO	KODAMA 269 GO	HIKARI 119 GO	HIKARI 27 GO	HIKARI 171 GO	KODAMA 235 GO	HIKARI 173 GO	KODAMA 261 GO	HIKARI 175 GO	KODAMA 265 GO	HIKARI 15 GO	HIKARI 193 GO	KODAMA 255 GO	HIKARI 177 GO	KODAMA 295 GO
TOKYO	15:24	15:40	15:48	16:00	16:12	16:16	16:24	16:40	16:48	16:52	17:00	17:12	17:16	17:24	17:40
SHINYOKOHAMA	-	15:58	-	-	-	16:34	-	16:58	-	17:10	-	-	17:34	-	17:58
ODAWARA	-	16:22	-	-	-	16:58	-	17:22	-	17:34	-	-	17:58	-	18:22
ATAMI	-	16:33	-	-	-	17:09	-	17:33	-	17:45	-	-	18:09	-	18:33
MISHIMA	-	16:47	-	-	-	17:23	-	17:47	-	17:59	-	-	18:23	-	18:47
SHIZUOKA	-	17:16	-	-	-	17:52	-	18:16	-	18:28	-	-	18:52	-	19:16
HAMAMATSU	-	17:43	-	-	-	18:19	-	18:43	-	18:55	-	-	19:19	-	19:43
TOYOHASHI	-	18:03	-	-	-	18:39	-	19:03	-	19:15	-	-	19:39	-	20:03
NAGOYA	17:27	18:32	17:51	18:03	18:15	19:08	18:27	19:32	18:51	19:44	19:03	19:15	20:08	19:27	20:32
GIFUHASHIMA	-	18:51	-	-	-	19:27	-	19:51	-	20:03	-	-	20:27	-	20:51
MAIBARA	-	19:09	-	-	-	19:45	-	20:09	-	20:21	-	19:41	20:45	-	21:09
KYOTO	18:17	19:36	18:41	18:53	19:05	20:12	19:17	20:36	19:41	20:48	19:53	20:08	21:12	20:17	21:36
SHINOSAKA	18:37	19:54	19:00	19:12	19:25	20:30	19:37	20:54	20:01	21:06	20:12	20:26	21:30	20:37	21:54
SHINKOBE	18:53		19:17	-	19:41		19:53		20:10		-			20:53	
NISHIAKASHI	19:05		-	-	19:53		20:05		20:35		-			21:05	
HIMEJI	19:20		19:45	-	20:08		20:20		20:49		-			21:20	
AIOI	19:31		-	-	20:19		20:31		21:00		-			21:31	
OKAYAMA	19:53		20:17	20:12	20:43		20:53		21:24		21:12			21:53	
SHINKURASHIKI			20:31	-	20:57				21:38		-				
FUKUYAMA			20:46	-	21:12				21:53		-				
MIHARA			21:01	-	21:27				22:08		-				
HIROSHIMA			21:29	21:10	21:53				22:34		22:10				
SHINIWAKUNI			21:48	-							-				
TOKUYAMA			22:08	-							-				
OGORI			22:28	21:58							-				
SHINSHIMONOSEKI			22:51	-							-				
KOKURA			23:04	22:29							23:24				
HAKATA			23:36	23:01							23:56				

	HIKARI 179 GO	KODAMA 291 GO	HIKARI 125 GO	HIKARI 195 GO	KODAMA 293 GO	HIKARI 183 GO	HIKARI 103 GO	KODAMA 297 GO	HIKARI 185 GO	HIKARI 29 GO	KODAMA 299 GO	HIKARI 93 GO	HIKARI 95 GO	HIKARI 99 GO
TOKYO	17:48	17:52	18:00	18:12	18:16	18:24	18:48	18:52	19:00	19:24	19:28	19:48	20:00	20:24
SHINYOKOHAMA	-	18:10	-	-	18:34	-	-	19:10	-	-	19:46	-	-	-
ODAWARA	-	18:34	-	-	18:58	-	-	19:34	-	-	20:10	-	-	-
ATAMI	-	18:45	-	-	19:09	-	-	19:45	-	-	20:21	-	-	-
MISHIMA	-	18:59	-	-	19:23	-	-	19:59	-	-	20:35	-	-	-
SHIZUOKA	-	19:23	-	-	19:52	-	-	20:26	-	-	21:04	-	-	-
HAMAMATSU	-	19:55	-	-	20:19	-	-	20:55	-	-	21:31	-	-	-
TOYOHASHI	-	20:15	-	-	20:39	-	-	21:15	-	-	21:51	-	-	-
NAGOYA	19:51	20:44	20:03	20:15	21:08	20:27	20:51	21:44	21:03	21:27	22:20	21:51	22:03	22:27
GIFUHASHIMA	-	21:03	-	-	21:27	-	-	22:03	-	-	22:39	-	-	-
MAIBARA	-	21:21	-	20:41	21:45	-	-	22:21	-	-	22:57	-	-	-
KYOTO	20:41	21:48	20:53	21:08	22:12	21:17	21:41	22:48	21:53	22:17	23:24	22:41	22:53	23:17
SHINOSAKA	21:01	22:06	21:13	21:26	22:30	21:37	21:58	23:06	22:13	22:34	23:42	22:58	23:10	23:34
SHINKOBE	21:17		21:29			21:53			22:29					
NISHIAKASHI	21:29		21:41			22:05			22:41					
HIMEJI	21:44		21:56			22:20			22:56					
AIOI	21:55		22:07			22:31			23:07					
OKAYAMA	22:18		22:31			22:53			23:29					
SHINKURASHIKI			22:45											
FUKUYAMA			23:00											
MIHARA			23:15											
HIROSHIMA			23:41											
SHINIWAKUNI														
TOKUYAMA														
OGORI														
SHINSHIMONOSEKI														
KOKURA														
HAKATA														

	KODAMA 228 GO	KODAMA 204 GO	HIKARI 94 GO	KODAMA 212 GO	KODAMA 200 GO	HIKARI 98 GO	KODAMA 202 GO	HIKARI 102 GO	KODAMA 206 GO	HIKARI 196 GO	KODAMA 208 GO	HIKARI 150 GO	KODAMA 210 GO	HIKARI 152 GO	KODAMA 214 GO
HAKATA												6:03		6:27	
KOKURA												6:25		6:49	
SHINSHIMONOSEKI												6:36		7:00	
OGORI												6:51		7:15	
TOKUYAMA												7:03		7:27	
SHINIWAKUNI			6:00		6:14	6:22	6:38	6:46	7:02	7:10	7:14	7:22	7:38	7:46	8:02
HIROSHIMA			6:19		6:33	6:41	6:57	7:05	7:21	7:23	7:33	7:41	7:57	8:05	8:21
MIHARA			-		6:59	-	7:23	-	7:47	-	7:59	-	8:23	-	8:47
FUKUYAMA			-		7:20	-	7:44	-	8:08	-	8:20	-	8:14	-	9:09
SHINKURASHIKI			7:09	7:17	7:36	73:10	8:00	7:55	8:24	8:19	8:36	8:31	9:00	8:55	9:24
OKAYAMA	6:20		-	7:43	8:07	-	8:31	-	8:55	-	9:07	-	9:31	-	9:55
AIOI	6:47		-	8:00	8:24	-	8:48	-	9:12	-	9:24	-	9:48	-	10:12
HIMEJI	7:04		-	8:30	8:54	-	9:18	-	9:42	-	9:54	-	10:18	-	10:42
NISHIAKASHI	7:33	8:00	-	8:59	9:23	-	9:47	-	10:11	-	10:23	-	10:47	-	11:11
SHINKOBE	7:57	8:25	-	9:10	9:34	-	9:58	-	10:22	-	10:34	-	10:58	-	11:22
SHINOSAKA	8:07	8:37	-	9:24	9:48	-	10:12	-	10:36	-	10:48	-	11:12	-	11:36
KYOTO	8:18	8:48	-	9:44	10:08	-	10:32	-	10:56	-	11:08	-	11:32	-	11:56
MAIBARA	8:39	9:08	9:10	10:04	10:28	9:32	10:52	9:56	11:16	10:20	11:28	10:32	11:52	10:56	12:16
GIFUHASHIMA	9:00	9:28													
NAGOYA															
TOYOHASHI															
HAMAMATSU															
SHIZUOKA															
MISHIMA															
ATAMI															
ODAWARA															
SHINYOKOHAMA															
TOKYO															

	HIKARI 190 GO	HIKARI 156 GO	HIKARI 158 GO	KODAMA 220 GO	HIKARI 160 GO	KODAMA 224 GO	HIKARI 162 GO	HIKARI 164 GO	HIKARI 2 GO	KODAMA 230 GO	HIKARI 192 GO	KODAMA 234 GO	HIKARI 168 GO	HIKARI 170 GO	KODAMA 239 GO
HAKATA															
KOKURA															
SHINSHIMONOSEKI															
OGORI															
TOKUYAMA															
SHINIWAKUNI															
HIROSHIMA			6:03					6:51	7:36						
NIHARA			6:30					7:18	-						
FUKUYAMA			6:44					7:32	-						
SHINKURASHIKI			7:00					7:48	-						
OKAYAMA		7:03	7:15		7:26		7:51	8:03	8:33				8:58	9:15	
AIOI		7:25	7:37		7:48		8:13	8:25	-				9:22	9:37	
HIMEJI		7:36	7:48		7:59		8:24	8:36	-				9:37	9:48	
NISHIAKASHI		7:51	8:03		8:14		8:39	8:51	-				9:52	10:03	
SHINKOBE		8:03	8:15		8:26		8:51	9:03	-				10:04	10:15	
SHINOSAKA	8:06	8:22	8:34	8:38	8:46	9:02	9:10	9:22	9:34	9:38	9:42	10:02	10:22	10:34	10:38
KYOTO	8:25	8:41	8:53	8:57	9:05	9:21	9:29	9:41	9:53	9:57	10:01	10:21	10:41	10:53	10:57
MAIBARA	8:52	-	-	9:23	-	9:47	-	-	-	10:23	10:28	10:47	-	-	11:23
GIFUHASHIMA	-	-	-	9:44	-	10:08	-	-	-	10:45	-	11:08	-	-	11:44
NAGOYA	9:19	9:31	9:43	10:00	9:55	10:24	10:19	10:31	10:43	11:00	10:55	11:24	11:31	11:43	12:00
TOYOHASHI	-	-	-	10:31	-	10:55	-	-	-	11:31	-	11:55	-	-	12:31
HAMAMATSU	-	-	-	10:48	-	11:12	-	-	-	11:48	-	12:12	-	-	12:48
SHIZUOKA	-	-	-	11:18	-	11:42	-	-	-	12:18	-	12:42	-	-	13:18
MISHIMA	-	-	-	11:47	-	12:11	-	-	-	12:47	-	13:11	-	-	13:47
ATAMI	-	-	-	11:58	-	12:22	-	-	-	12:58	-	13:22	-	-	13:58
ODAWARA	-	-	-	12:12	-	12:36	-	-	-	13:12	-	13:36	-	-	14:12
SHINYOKOHAMA	-	-	-	12:32	-	12:56	-	-	-	13:32	-	13:56	-	-	14:32
TOKYO	11:20	11:32	11:44	12:52	11:56	13:16	12:20	12:32	12:44	13:52	12:56	14:16	13:32	13:44	14:52

	HIKARI 100 GO	HIKARI 20 GO	KODAMA 232 GO	HIKARI 104 GO	HIKARI 4 GO	KODAMA 254 GO	HIKARI 172 GO	KODAMA 256 GO	HIKARI 106 GO	HIKARI 6 GO	KODAMA 250 GO	HIKARI 174 GO	KODAMA 252 GO	HIKARI 108 GO	HIKARI 8 GO
HAKATA	6:06	6:19		6:44	7:24				7:44	8:24				8:44	9:24
KOKURA	6:39	6:52		7:17	7:57				8:17	8:57				9:17	9:57
SHINSHIMONOSEKI	6:51	-		7:29	-				8:29	-				9:29	-
OGORI	7:14	7:21		7:52	-				8:52	-				9:52	-
TOKUYAMA	7:40	-		8:12	-				9:12	-				10:12	-
SHINIWAKUNI	8:00	-		8:32	-				9:32	-				10:32	-
HIROSHIMA	8:21	8:12		8:53	9:12				9:53	10:12				10:53	11:12
MIHARA	8:48	-		9:20	-				10:20	-				11:20	-
FUKUYAMA	9:02	-		9:34	-				10:34	-				11:34	-
SHINKURASHIKI	9:18	-		9:50	-				10:50	-				11:50	-
OKAYAMA	9:33	9:09		10:05	10:09		10:27		11:05	11:09		11:27		12:05	12:09
AIOI	-	-		-	-		10:49		-	-		11:49		-	-
HIMEJI	10:03	-		10:37	-		11:00		11:37	-		12:00		12:37	-
NISHIAKASHI	-	-		-	-		11:15		-	-		12:15		-	-
SHINKOBE	10:28	-		11:03	-		11:27		12:03	-		12:27		13:03	-
SHINOSAKA	10:46	10:10	11:02	11:22	11:10	11:38	11:46	12:02	12:22	12:10	12:38	12:46	13:02	13:22	13:10
KYOTO	11:05	10:29	11:21	11:41	11:29	11:57	12:05	12:21	12:41	12:29	12:57	13:05	13:21	13:41	13:29
MAIBARA	-	-	11:47	-	-	12:23	-	12:47	-	-	13:23	-	13:47	-	-
GIFUHASHIMA	-	-	12:08	-	-	12:44	-	13:08	-	-	13:44	-	14:08	-	-
NAGOYA	11:55	11:19	12:24	12:31	12:19	13:00	12:55	13:24	13:31	13:19	14:00	13:55	14:24	14:31	14:19
TOYOHASHI	-	-	12:55	-	-	13:31	-	13:55	-	-	14:31	-	14:55	-	-
HAMAMATSU	-	-	13:12	-	-	13:48	-	14:12	-	-	14:48	-	15:12	-	-
SHIZUOKA	-	-	13:42	-	-	14:18	-	14:42	-	-	15:18	-	15:42	-	-
MISHIMA	-	-	14:11	-	-	14:47	-	15:11	-	-	15:47	-	16:11	-	-
ATAMI	-	-	14:22	-	-	14:58	-	15:22	-	-	15:58	-	16:22	-	-
ODAWARA	-	-	14:36	-	-	15:12	-	15:36	-	-	16:12	-	16:36	-	-
SHINYOKOHAMA	-	-	14:56	-	-	15:32	-	15:56	-	-	16:32	-	16:56	-	-
TOKYO	13:56	13:20	15:16	14:32	14:20	15:52	14:56	16:16	15:32	15:20	16:52	15:56	17:16	16:32	16:20

	KODAMA 258 GO	HIKARI 176 GO	KODAMA 260 GO	HIKARI 110 GO	HIKARI 22 GO	KODAMA 266 GO	HIKARI 178 GO	KODAMA 268 GO	HIKARI 120 GO	HIKARI 10 GO	HIKARI 90 GO	KODAMA 264 GO	HIKARI 180 GO	KODAMA 236 GO	KODAMA 262 GO
HAKATA				9:44	10:19				10:54	11:24					
KOKURA				10:17	10:52				11:27	11:57					
SHINSHIMONOSEKI				10:29	-				11:39	-					
OGORI				10:52	11:21				12:02	-					
TOKUYAMA				11:12	-				12:22	-					
SHINIWAKUNI				11:32	-				12:42	-					
HIROSHIMA				11:53	12:12				13:03	13:12					
MIHARA				12:20	-				13:30	-					
FUKUYAMA				12:34	-				13:47	-					
SHINKURASHIKI				12:50	-				14:05	-					
OKAYAMA		12:27		13:05	13:09		13:27		14:21	14:09			14:27		
AIOI		12:49		-	-		13:40		-	-			14:49		
HIMEJI		13:00		13:37	-		14:00		-	-			15:00		
NISHIAKASHI		13:15		-	-		14:15		-	-			15:15		
SHINKOBE		13:27		14:03	-		14:27		-	-			15:27		
SHINOSAKA	13:38	13:45	14:02	14:22	14:10	14:33	14:46	15:02	15:22	15:10	15:34	15:38	15:46	16:02	16:14
KYOTO	13:57	14:05	14:21	14:41	14:29	14:57	15:05	15:21	15:41	15:29	15:53	15:57	16:05	16:21	16:33
MAIBARA	14:23	-	14:47	-	-	15:23	-	15:47	-	-	-	16:23	-	16:47	16:59
GIFUHASHIMA	14:44	-	15:08	-	-	15:44	-	16:08	-	-	-	16:44	-	17:08	17:20
NAGOYA	15:00	14:55	15:24	15:31	15:19	16:00	15:55	16:24	16:31	16:19	16:43	17:00	16:55	17:24	17:36
TOYOHASHI	15:31	-	15:55	-	-	16:31	-	16:55	-	-	-	17:31	-	17:55	18:07
HAMAMATSU	15:48	-	16:12	-	-	16:48	-	17:12	-	-	-	17:48	-	18:12	18:24
SHIZUOKA	16:18	-	16:42	-	-	17:18	-	17:42	-	-	-	18:18	-	18:42	18:54
MISHIMA	16:47	-	17:11	-	-	17:47	-	18:11	-	-	-	18:47	-	19:11	19:23
ATAMI	16:58	-	17:22	-	-	17:58	-	18:22	-	-	-	18:58	-	19:22	19:34
ODAWARA	17:12	-	17:36	-	-	18:12	-	18:36	-	-	-	19:12	-	19:36	19:48
SHINYOKOHAMA	17:32	-	17:56	-	-	18:32	-	18:56	-	-	-	19:32	-	19:56	20:08
TOKYO	17:52	16:56	18:16	17:32	17:20	18:52	17:56	19:16	18:32	18:20	18:44	19:52	18:56	20:16	20:28

	HIKARI 112 GO	HIKARI 12 GO	HIKARI 154 GO	KODAMA 292 GO	HIKARI 182 GO	KODAMA 284 GO	HIKARI 114 GO	HIKARI 24 GO	HIKARI 166 GO	KODAMA 288 GO	HIKARI 184 GO	KODAMA 290 GO	HIKARI 116 GO	HIKARI 14 GO	KODAMA 294 GO
HAKATA	11:44	12:24					12:44	13:19					13:44	14:24	
KOKURA	12:17	12:57					13:17	13:52					14:17	14:57	
SHINSHIMONOSEKI	12:29	-					13:29	-					14:29	-	
OGORI	12:52	-					13:52	14:21					14:52	-	
TOKUYAMA	13:12	-					14:12	-					15:12	-	
SHINIWAKUNI	13:32	-					14:32	-					15:32	-	
HIROSHIMA	13:53	14:12					14:53	15:12					15:53	16:12	
MIHARA	14:20	-					15:20	-					16:20	-	
FUKUYAMA	14:34	-					15:34	-					16:34	-	
SHINKURASHIKI	14:50	-					15:50	-					16:50	-	
OKAYAMA	15:05	15:09	15:23		15:27		16:05	16:09	16:23		16:27		17:05	17:09	
AIOI	-	-	-		15:49		-	-	-		16:49		-	-	
HIMEJI	15:37	-	15:53		16:00		16:37	-	16:53		17:00		17:37	-	
NISHIAKASHI	-	-	-		16:15		-	-	-		17:15		-	-	
SHINKOBE	16:03	-	16:16		16:27		17:03	-	17:16		17:27		18:03	-	
SHINOSAKA	16:22	16:10	16:34	16:38	16:46	17:02	17:22	17:10	17:34	17:38	17:46	18:02	18:22	18:10	18:38
KYOTO	16:41	16:29	16:53	16:58	17:06	17:21	17:41	17:29	17:53	17:57	18:05	18:21	18:41	18:29	18:57
MAIBARA	-	-	-	17:24	-	17:47	-	-	-	18:23	-	18:47	-	-	19:23
GIFUHASHIMA	-	-	-	17:45	-	18:08	-	-	-	18:44	-	19:08	-	-	19:44
NAGOYA	17:31	17:19	17:43	18:00	17:55	18:24	18:31	18:19	18:43	19:00	18:55	19:24	19:31	19:19	20:00
TOYOHASHI	-	-	-	18:31	-	18:55	-	-	-	19:31	-	19:55	-	-	20:31
HAMAMATSU	-	-	-	18:48	-	19:12	-	-	-	19:48	-	20:12	-	-	20:48
SHIZUOKA	-	-	-	19:18	-	19:42	-	-	-	20:18	-	20:42	-	-	21:18
MISHIMA	-	-	-	19:47	-	20:11	-	-	-	20:47	-	21:11	-	-	21:47
ATAMI	-	-	-	19:58	-	20:22	-	-	-	20:58	-	21:22	-	-	21:58
ODAWARA	-	-	-	20:12	-	20:36	-	-	-	21:12	-	21:36	-	-	22:12
SHINYOKOHAMA	-	-	-	20:32	-	20:56	-	-	-	21:32	-	21:56	-	-	22:32
TOKYO	19:32	19:20	19:44	20:52	19:56	21:16	20:32	20:20	20:44	21:52	20:56	22:16	21:32	21:20	22:52

	HIKARI 186 GO	KODAMA 296 GO	HIKARI 122 GO	HIKARI 26 GO	KODAMA 298 GO	HIKARI 96 GO	HIKARI 183 GO	KODAMA 216 GO	HIKARI 194 GO	HIKARI 118 GO	HIKARI 16 GO	KODAMA 218 GO	HIKARI 18 GO	KODAMA 222 GO	HIKARI 28 GO
HAKATA			14:54	15:19				15:36		15:54	16:24	16:36	17:24	17:36	18:24
KOKURA			15:27	15:52				16:09		16:27	16:57	17:09	17:57	18:09	18:57
SHINSHIMONOSEKI			15:39	-				16:21		16:39	-	17:21	-	18:21	19:09
OGORI			16:02	16:21				16:44		17:02	-	17:44	-	18:44	-
TOKUYAMA			16:22	-				17:04		17:22	-	18:04	-	19:04	19:46
SHINIWAKUNI			16:42	-				17:24		17:42	-	18:24	-	19:24	-
HIROSHIMA			17:03	17:12				17:45		18:03	18:12	18:45	19:12	19:45	20:22
MIHARA			17:30	-				18:12		18:30	-	19:12	-	20:12	-
FUKUYAMA			17:47	-				18:26		18:47	-	19:26	-	20:26	-
SHINKURASHIKI			18:05	-				18:42		19:06	-	19:42	-	20:42	-
OKAYAMA	17:27		18:21	18:09			18:27	18:57		19:23	19:09	19:57	20:09	20:57	21:19
AIOI	17:49		-	-			18:49	19:19		-	-	20:19	-	21:19	-
HIMEJI	18:00		-	-			19:00	19:30		19:53	-	20:30	-	21:30	-
NISHIAKASHI	18:15		-	-			19:15	19:48		-	-	20:48	-	21:45	-
SHINKOBE	18:27		-	-			19:27	20:00		20:16	-	21:00	-	21:57	-
SHINOSAKA	18:46	19:02	19:22	19:10	19:26	19:34	19:46	20:16	20:18	20:34	20:10	21:16	21:10	22:13	22:18
KYOTO	19:05	19:21	19:41	19:29	19:45	19:53	20:05		20:37	20:53	20:29		21:29		
MAIBARA	-	19:47	-	-	20:11	-	-		21:04	-	-		-		
GIFUHASHIMA	-	20:08	-	-	20:32	-	-		-	-	-		-		
NAGOYA	19:55	20:24	20:31	20:19	20:48	20:43	20:55		21:31	21:43	21:19		22:17		
TOYOHASHI	-	20:55	-	-	21:19	-	-		-	-	-				
HAMAMATSU	-	21:12	-	-	21:36	-	-		-	-	-				
SHIZUOKA	-	21:42	-	-	22:06	-	-		-	-	-				
MISHIMA	-	22:11	-	-	22:35	-	-		-	-	-				
ATAMI	-	22:22	-	-	22:46	-	-		-	-	-				
ODAWARA	-	22:36	-	-	23:00	-	-		-	-	-				
SHINYOKOHAMA	-	22:56	-	-	23:20	-	-		-	-	-				
TOKYO	21:56	23:16	22:32	22:20	23:40	22:44	22:56		23:32	23:44	23:20				

	KODAMA 226 GO	HIKARI 92 GO
HAKATA	18:49	19:24
KOKURA	19:22	19:57
SHINSHIMONOSEKI	19:34	20:09
OGORI	19:57	-
TOKUYAMA	20:17	20:46
SHINIWAKUNI	20:37	-
HIROSHIMA	20:58	21:22
MIHARA	21:25	-
FUKUYAMA	21:40	-
SHINKURASHIKI	21:57	-
OKAYAMA	22:13	22:19
AIOI	22:40	-
HIMEJI	22:52	-
NISHIAKASHI	23:06	-
SHINKOBE	23:18	-
SHINOSAKA	23:34	23:18
KYOTO		
MAIBARA		
GIFUHASHIMA		
NAGOYA		
TOYOHASHI		
HAMAMATSU		
SHIZUOKA		
MISHIMA		
ATAMI		
ODAWARA		
SHINYOKOHAMA		
TOKYO		

付録 5. 会話音声認識システム（第２次）の発声リスト

(1) ～ (20)　認識実験に使用

(1) ～ (40)　質問回答実験に使用

(1)　１日の，新大阪から，博多までの，６時２分発の，ひかり１９号の，指定券を，９枚，お願いします。

(2)　新神戸発の，６時２２分の，グリーン券を，小倉まで，８枚。

(3)　３日の，こだま２１３号で，普通席を，西明石から，新下関まで，７枚，予約します。

(4)　岡山から，広島への，４日の，ひかり９１号の，指定席を，６枚。

(5)　姫路駅より，小郡駅までの，グリーン券で，５日の，こだま２１７号を，５枚。

(6)　６日の，ひかり１７号で，京都から，博多までの，指定を，４枚。

(7)　相生発の，９時３０分の，徳山までの，グリーン券は，ありますか。

(8)　東京から，名古屋までの，６時０分発の，グリーン券で，８日のは，ありますか。

(9)　９日ので，新横浜発の，６時２２分の，こだま２０１号の，米原ゆきの，指定券は，ありませんか。

(10)　新倉敷から，１０時４９分発の，新岩国までの，指定を，１枚。

(11)　ひかり１５１号で，１１号のを，名古屋から，相生まで，２枚，グリーン券を，予約します。

(12)　１２日で，小田原発の，７時１０分の，岐阜羽島までの，３枚の，指定券を，申し込みます。

(13)　ひかり１２１号で，京都から，三原までの，指定席で，１３日の，９時４１分発は，ありませんか。

(14)　１４日の，熱海から，７時４５分発の，こだま２０７号で，５枚の，グリーン券は。

(15)　１５日の，ひかり１号で，東京から，小倉までの，普通席を，６枚。

(16)　三島から，豊橋への，指定を，７枚。

(17)　１７日，ひかり１５３号，名古屋発の，９時２７分の，姫路への，グリーン券を，８枚。

(18)　１８日ので，新大阪駅から，博多駅までの，ひかり１０５号の，グリーンを，９枚。

(19)　静岡から，京都までの，指定券で，１９日の，９時１６分発のは，ありますか。

⑳　２０日の，ひかり１９１号で，東京から，７時４８分発，名古屋いきの，指定を，８枚。

㉑　２１日，ひかり３号，京都から，広島まで。

㉒　浜松から，米原まで，２２日の，こだま２１５号の，指定を，６枚。

㉓　ひかり１５３号で，１０時５３分発のを，相生まで，５枚。

㉔　２４日の，ひかり１０５号で，姫路から，１１時２８分発の，指定を，４枚，徳山まで。

㉕　２５日で，こだま２１１号で，豊橋から，新大阪まで，グリーンを，３枚。

㉖　２６日の，７時４８分，東京発の，ひかり１９１号の，普通を，２枚，米原まで。

㉗　２７日の，指定券で，岡山から，小倉まで，１２時１２分のを。

㉘　新横浜から，浜松への，２８日の，こだま２１５号の，グリーンを，１枚。

㉙　２９日の，１０時２７分，名古屋発の，西明石ゆきの，グリーンを。

㉚　熱海から，岐阜羽島駅まで，指定券を，３枚，３０日の，９時３３分発の，こだま２１９号で。

㉛　３１日の，ひかり１０７号で，新神戸駅から，小郡駅への，グリーンを，４枚。

㉜　９時０分発，東京からで，博多までの，１日の，グリーンを，５枚，予約します。

㉝　２日の，こだま２２３号で，三島から，豊橋まで，６枚，お願い致します。

㉞　名古屋から，新大阪まで，３日の，指定は，ありませんか。

㉟　こだま２２５号で，静岡より，新大阪までの，指定で，８枚，４日のは，ありませんか。

㊱　５日の，新倉敷から，小郡まで，１４時３１分発の，ひかり１０９号は，ありませんか。

㊲　６日ので，京都より，小倉いきの，グリーン券で，１２時５３分発の，ひかり５号で，９枚。

㊳　７日の，ひかり１２７号で，東京から，岡山まで，グリーンを，８枚。

㊴　浜松から，新大阪まで，８日の，こだま２２９号で，指定券を，７枚，お願いします。

㊵　東京から，西明石ゆきの，９日の，指定券は，ありますか。

付録 6. 単語音声認識実験における単語辞書

図A 6 − 1　　月

```
 1  1   2    9  I  *  CH I  G  A  *  TS  U
             24  4  1  2  4  2  4  1  2   4
             82  8  7  7 12  6 14  6  8  14
 2  1   0    9  I  *  CH I  G  A  *  TS  U'
             20  4  1  2  4  2  4  1  2   0
             78  8  7  7 12  6 14  6  8  10
 3  1  -2    6  I  *  CH I  G  A
             17  4  1  2  4  2  4
             54  8  7  7 12  6 14
 4  2   0    7  NI I- G  A  *  TS U
             17  2  2  2  4  1  2  4
             64  6 10  6 14  6  8 14
 5  2   0    7  NI I- G  A  *  TS U'
             13  2  2  2  4  1  2  0
             60  6 10  6 14  6  8 10
 6  2  -5    4  NI I- G  A
             10  2  2  2  4
             36  6 10  6 14
 7  3   0    8  S  A  N- G  A  *  TS U
             24  2  5  4  2  4  1  2  4
             86  8 14 16  6 14  6  8 14
 8  3   0    8  S  A  N- G  A  *  TS U'
             20  2  5  4  2  4  1  2  0
             82  8 14 16  6 14  6  8 10
 9  3  -8    5  S  A  N- G  A
             17  2  5  4  2  4
             58  8 14 16  6 14
10  4   0    7  SH I  G  A  *  TS U
             15  0  2  2  4  1  2  4
             64  8  8  6 14  6  8 14
11  4   0    7  SH I  G  A  *  TS U'
             11  0  2  2  4  1  2  0
             60  8  8  6 14  6  8 10
12  4 -11    4  SH I  G  A
              8  0  2  2  4
             36  8  8  6 14
13  5   0    7  G  0  G  A  *  TS U
             18  2  3  2  4  1  2  4
             68  6 14  6 14  6  8 14
14  5   0    7  G  0  G  A  *  TS U'
             14  2  3  2  4  1  2  0
             64  6 14  6 14  6  8 10
15  5 -14    4  G  0  G  A
             11  2  3  2  4
             40  6 14  6 14
16  6   0   10  R  0  *  KU U- G  A  *  TS U
             22  2  4  0  1  2  2  4  1  2  4
             85  8 12  4  3 10  6 14  6  8 14
```

```
17  6   0   10  R  0  *  KU U- G  A  *  TS U'
            18  2  4  0  1  2  2  4  1  2  0
            81  8 12  4  3 10  6 14  6  8 10
18  6 -17    7  R  0  *  KU U- G  A
            15  2  4  0  1  2  2  4
            57  8 12  4  3 10  6 14
19  7   0   10  SH I  *  CH I  G  A  *  TS U
            24  2  2  1  2  4  2  4  1  2  4
            91  8  9  7  7 12  6 14  6  8 14
20  7   0   10  SH I  *  CH I  G  A  *  TS U'
            20  2  2  1  2  4  2  4  1  2  0
            87  8  9  7  7 12  6 14  6  8 10
21  7 -20    7  SH I  *  CH I  G  A
            17  2  2  1  2  4  2  4
            63  8  9  7  7 12  6 14
22  8   0   11  H  A  AI *  CH I  G  A  *  TS U
            24  2  2  0  1  2  4  2  4  1  2  4
            92  6  7  5  7  7 12  6 14  6  8 14
23  8   0   11  H  A  AI *  CH I  G  A  *  TS U'
            20  2  2  0  1  2  4  2  4  1  2  0
            88  6  7  5  7  7 12  6 14  6  8 10
24  8 -23    8  H  A  AI *  CH I  G  A
            17  2  2  0  1  2  4  2  4
            64  6  7  5  7  7 12  6 14
25  9   0    7  KU U- G  A  *  TS U
            16  1  2  2  4  1  2  4
            61  3 10  6 14  6  8 14
26  9   0    7  KU U- G  A  *  TS U'
            12  1  2  2  4  1  2  0
            57  3 10  6 14  6  8 10
27  9 -26    4  KU U- G  A
             9  1  2  2  4
            33  3 10  6 14
28 10   0    9  Z  I  IU U  G  A  *  TS U
            24  2  2  3  4  2  4  1  2  4
            90  6 10 12 14  6 14  6  8 14
29 10   0    9  Z  I  IU U  G  A  *  TS U'
            20  2  2  3  4  2  4  1  2  0
            86  6 10 12 14  6 14  6  8 10
30 10 -29    6  Z  I  IU U  G  A
            17  2  2  3  4  2  4
            62  6 10 12 14  6 14
31 11   0   14  Z  I  IU U  UI I  *  CH I  G  A  *  TS U
            33  2  2  3  2  2  3  1  2  3  2  4  1  2  4
           132  6 10 12 10 10 12  7  7 10  6 14  6  8 14
32 11   0   14  Z  I  IU U  UI I  *  CH I  G  A  *  TS U'
            29  2  2  3  2  2  3  1  2  3  2  4  1  2  0
           128  6 10 12 10 10 12  7  7 10  6 14  6  8 10
33 11 -33   11  Z  I  IU U  UI I  *  CH I  G  A
            26  2  2  3  2  2  3  1  2  3  2  4
           104  6 10 12 10 10 12  7  7 10  6 14
34 12   0   11  Z  I  IU U  NI I- G  A  *  TS U
            29  2  2  3  4  3  2  2  4  1  2  4
           108  6 10 12 14  8 10  6 14  6  8 14
```

```
35 12   0   11  Z  I  IU U  NI I- G  A  *  TS U'
            25  2  2  3  4  3  2  2  4  1  2  0
           104  6 10 12 14  8 10  6 14  6  8 10
36 12 -35    8  Z  I  IU U  NI I- G  A
            22  2  2  3  4  3  2  2  4
            80  6 10 12 14  8 10  6 14
37 13   0    2  I  E
            12  8  4
            39 24 15
```

図A 6－2　　日

```
 1  1   0   10  TS UI I  *  T  A  AI *  CH I
            24  3  3  3  1  1  4  0  1  2  6
            84  8  8 14  8  3 10  6  4  5 18
 2  1   0   10  TS UI I  *  T  A  AI *  CH I'
            18  3  3  3  1  1  4  0  1  2  0
            76  8  8 14  8  3 10  6  4  5 10
 3  1  -2    7  TS UI I  *  T  A  AI
            15  3  3  3  1  1  4  0
            57  8  8 14  8  3 10  6
 4  2   0    8  F  U  *  TS U  *  KA A
            23  2  2  1  2  8  1  1  6
            67  6  6  6  7 13  8  3 18
 5  2   0    8  F  U' *  TS U  *  KA A
            16  2  0  1  2  3  1  1  6
            60  6  4  6  7  8  8  3 18
 6  2   5    5  TS U  *  KA A
            14  2  3  1  1  7
            44  7  8  8  3 18
 7  2   0    8  F  U  *  TS U' *  KA A
            15  2  2  1  2  0  1  1  6
            58  6  6  6  7  4  8  3 18
 8  2   0    8  F  U' *  TS U' *  KA A
            13  2  0  1  2  0  1  1  6
            56  6  4  6  7  4  8  3 18
 9  2   8    5  TS U' *  KA A
            10  2  0  1  1  6
            40  7  4  8  3 18
10  3   0    5  M  I  *  KA A
            23  2  4 10  1  6
            59  7 12 19  3 18
11  4   0    6  I  I0 0  *  KA A
            22  2  3  2  8  1  6
            63  6  7 10 19  3 18
12  5   0    7  I  *  TS U  *  KA A
            18  4  1  2  3  1  1  6
            62 10  8  7  8  8  3 18
13  5   0    7  I' *  TS U  *  KA A
            17  3  1  2  3  1  1  6
            59  7  8  7  8  8  3 18
14  5  13    5  TS U  *  KA A
            13  2  3  1  1  6
            44  7  8  8  3 18
```

```
15  5   0    7  I  *  TS U' *  KA A
            15  4  1  2  0  1  1  6
            62 10  8  7  8  8  3 18
16  5   0    7  I' *  TS U' *  KA A
            14  3  1  2  0  1  1  6
            55  7  8  7  4  8  3 18
17  5  16    5  TS U' *  KA A
            10  2  0  1  1  6
            44  7  8  8  3 18
18  6   0    6  M  UI I  *  KA A
            18  3  2  5  1  1  6
            60  9  7 15  8  3 18
19  7   0    7  N  A  N  0  *  KA A
            22  2  4  2  4  1  1  8
            65  6 11  6 13  6  3 20
20  8   0    6  1  IO 0  *  KA A
            20  2  2  6  1  1  8
            61  6  7 17  8  3 20
21  9   0   10  KO 0  *  KO 0  N  0  *  KA A
            22  2  2  1  2  2  2  3  1  1  6
            83  4 12  6  4 12  7 11  6  3 18
22  9  21    7  KO 0  N  0  *  KA A
            17  2  2  2  3  1  1  6
            61  4 12  7 11  6  3 18
23 10   0    5  T  0  *  KA A
            21  1 10  1  1  8
            55  3 20  7  3 22
24 11   0   14  Z  I  IU U  UI I  *  CH I  NI I- *  CH I
            33  2  2  3  2  2  3  1  2  3  2  4  1  2  4
           130  7 10 12 10  7 10  7  7  9  7 11  8  7 18
25 11   0   14  Z  I  IU U  UI I  *  CH I  NI I- *  CH I'
            29  2  2  3  2  2  3  1  2  3  2  4  1  2  0
           122  7 10 12 10  7 10  7  7  9  7 11  8  7 10
26 11 -25   11  Z  I  IU U  UI I  *  CH I  NI I-
            26  2  2  3  2  2  3  1  2  3  2  4
            97  7 10 12 10  7 10  7  7  9  7 11
27 12   0   11  Z  I  IU U  NI I- NI I- *  CH I
            28  2  2  3  2  2  4  2  4  1  2  4
           110  7 10 12 12  7 11  7 11  8  7 18
28 12   0   11  Z  I  IU U  NI I- NI I- *  CH I'
            32  2  2  3  2  2  4  2  4  1  2  8
           102  7 10 12 12  7 11  7 11  8  7 10
29 12 -28    8  Z  I  IU U  NI I- NI I-
            21  2  2  3  2  2  4  2  4
            77  7 10 12 12  7 11  7 11
30 13   0   12  Z  I  IU U  S  A  N  NI I- *  CH I
            35  2  2  3  2  4  6  3  2  4  1  2  4
           127  7 10 12 12 10 13 12  7 11  8  7 18
31 13   0   12  Z  I  IU U  S  A  N  NI I- *  CH I'
            35  2  2  3  2  4  6  3  2  4  1  2  4
           127  7 10 12 12 10 13 12  7 11  8  7 18
32 13 -31    9  Z  I  IU U  S  A  N  NI I-
            28  2  2  3  2  4  6  3  2  4
            94  7 10 12 12 10 13 12  7 11
```

```
33 14   0   11  Z  I  IU U  UI I  IO 0  *  KA A
            35  2  2  3  2  2  2  2  3  8  1  8
           108  7 10 12 10  7  5  7  8 17  3 22
34 15   0   12  Z  I  IU U  G  0  0I NI I- *  CH I
            27  2  2  3  2  2  3  0  2  4  1  2  4
           107  7 10 12 12  4  7  4  7 11  8  7 18
35 15   0   12  Z  I  IU U  G  0  0I NI I- *  CH I'
            23  2  2  3  2  2  3  0  2  4  1  2  0
            99  7 10 12 12  4  7  4  7 11  8  7 10
36 15 -35    9  Z  I  IU U  G  0  0I NI I-
            20  2  2  3  2  2  3  0  2  4
            74  7 10 12 12  4  7  4  7 11
37 16   0   14  Z  I  IU U  R  0  *  KU U- NI I- *  CH I
            32  2  2  3  2  2  4  1  1  2  2  4  1  2  4
           120  7 10 12 12  5 10  4  3  6  7 11  8  7 18
38 16   0   14  Z  I  IU U  R  0  *  KU U- NI I- *  CH I'
            28  2  2  3  2  2  4  1  1  2  2  4  1  2  0
           112  7 10 12 12  5 10  4  3  6  7 11  8  7 10
39 16 -38   11  Z  I  IU U  R  0  *  KU U- NI I-
            25  2  2  3  2  2  4  1  1  2  2  4
            87  7 10 12 12  5 10  4  3  6  7 11
40 17   0   15  Z  I  IU U  UI SH I  *  CH I  NI I- *  CH I
            34  2  2  3  2  2  3  2  1  2  2  2  4  1  2  4
           128  7 10 12 10  5  6  5  7  7  8  7 11  8  7 18
41 17   0   15  Z  I  IU U  UI SH I  *  CH I  NI I- *  CH I'
            30  2  2  3  2  2  3  2  1  2  2  2  4  1  2  0
           120  7 10 12 10  5  6  5  7  7  8  7 11  8  7 10
42 17 -41   12  Z  I  IU U  UI SH I  *  CH I  NI I-
            27  2  2  3  2  2  3  2  1  2  2  2  4
            95  7 10 12 10  5  6  5  7  7  8  7 11
43 17   0   15  Z  I  IU U  UI SH I' *  CH I  NI I- *  CH I
            31  2  2  3  2  2  3  0  0  2  2  2  4  1  2  4
           127  7 10 12 10  5  6  4  7  7  8  7 11  8  7 18
44 17   0   15  Z  I  IU U  UI SH I' *  CH I  NI I- *  CH I'
            27  2  2  3  2  2  3  0  0  2  2  2  4  1  2  0
           119  7 10 12 10  5  6  4  7  7  8  7 11  8  7 10
45 17 -44   12  Z  I  IU U  UI SH I' *  CH I  NI I-
            24  2  2  3  2  2  3  0  0  2  2  2  4
            94  7 10 12 10  5  6  4  7  7  8  7 11
46 18   0   15  Z  I  IU U  H  A  AI *  CH I  NI I- *  CH I
            31  2  2  3  2  2  2  0  1  2  2  2  4  1  2  4
           132  7 10 12 12  6  7  5  7  7  8  7 11  8  7 18
47 18   0   15  Z  I  IU U  H  A  AI *  CH I  NI I- *  CH I'
            27  2  2  3  2  2  2  0  1  2  2  2  4  1  2  0
           124  7 10 12 12  6  7  5  7  7  8  7 11  8  7 10
48 18 -47   12  Z  I  IU U  H  A  AI *  CH I  NI I-
            24  2  2  3  2  2  2  0  1  2  2  2  4
            99  7 10 12 12  6  7  5  7  7  8  7 11
49 19   0   12  Z  I  IU U  *  KU U- NI I- *  CH I
            26  2  2  3  2  1  1  2  2  4  1  2  4
           107  7 10 12 12  6  3  6  7 11  8  7 18
50 19   0   12  Z  I  IU U  *  KU U- NI I- *  CH I'
            22  2  2  3  2  1  1  2  2  4  1  2  0
            99  7 10 12 12  6  3  6  7 11  8  7 10
```

```
51 19 -50    9  Z  I  IU U  *  KU U- NI I-
            19  2  2  3  2  1  1  2  2  4
            74  & 10 12 12  6  3  6  7 11
52 20   0    8  H  A  *  TS U  *  KA A
            20  2  4  1  2  3  1  1  6
            66  6 10  6  7  8  8  3 18
53 20   0    8  H  A  *  TS U' *  KA A
            17  2  4  1  2  0  1  1  6
            62  6 10  6  7  4  8  3 18
54 21   0   16  NI I- Z- I  IU U  UI I  *  CH I  NI I- *  CH I
            41  2  6  2  2  3  2  2  3  1  2  3  2  4  1  2  4
           141  7 11  3  7 12 10  7 10  7  7  9  7 11  8  7 18
55 21   0   16  NI I- Z- I  IU U  UI I  *  CH I  NI I- *  CH I'
            37  2  6  2  2  3  2  2  3  1  2  3  2  4  1  2  0
           133  7 11  3  7 12 10  7 10  7  7  9  7 11  8  7 10
56 21 -55   13  NI I- Z- I  IU U  UI I  *  CH I  NI I-
            34  2  6  2  2  3  2  2  3  1  2  3  2  4
           108  7 11  3  7 12 10  7 10  7  7  9  7 11
57 22   0   13  NI I- Z- I  IU U  NI I- NI I- *  CH I
            36  2  6  2  2  3  2  2  4  2  4  1  2  4
           121  7 11  3  7 12 12  7 11  7 11  8  7 18
58 22   0   13  NI I- Z- I  IU U  NI I- NI I- *  CH I'
            32  2  6  2  2  3  2  2  4  2  4  1  2  0
           113  7 11  3  7 12 12  7 11  7 11  8  7 10
59 22 -58   10  NI I- Z- I  IU U  NI I- NI I-
            29  2  6  2  2  3  2  2  4  2  4
            88  7 11  3  7 12 12  7 11  7 11
60 23   0   14  NI I- Z- I  IU U  S  A  N  NI I- *  CH I
            43  2  6  2  2  3  2  4  6  3  2  4  1  2  4
           138  & 11  3  7 12 12 10 13 12  7 11  8  7 18
61 23   0   14  NI I- Z- I  IU U  S  A  N  NI I- *  CH I'
            39  2  6  2  2  3  2  4  6  3  2  4  1  2  0
           130  7 11  3  7 12 12 10 13 12  7 11  8  7 10
62 23 -61   11  NI I- Z- I  IU U  S  A  N  NI I-
            36  2  6  2  2  3  2  4  6  3  2  4
           105  7 11  3  7 12 12 10 13 12  7 11
63 24   0   13  NI I- Z- I  IU U  UI I  IO O  *  KA A
            43  2  6  2  2  3  2  2  2  2  3  8  1  8
           119  7 11  3  7 12 10  7  5  7  8 17  3 22
64 25   0   14  NI I- Z- I  IU U  G  O  OI NI I- *  CH I
            35  2  6  2  2  3  2  2  3  0  2  4  1  2  4
           118  7 11  3  7 12 12  4  7  4  7 11  8  7 18
65 25   0   14  NI I- Z- I  IU U  G  O  OI NI I- *  CH I'
            31  2  6  2  2  3  2  2  3  0  2  4  1  2  0
           110  7 11  3  7 12 12  4  7  4  7 11  8  7 10
66 25 -65   11  NI I- Z- I  IU U  G  O  OI NI I-
            28  2  6  2  2  3  2  2  3  0  2  4
            85  7 11  3  7 12 12  4  7  4  7 11
67 26   0   16  NI I- Z- I  IU U  R  O  *  KU U- NI I- *  CH I
            40  2  6  2  2  3  2  2  4  1  1  2  2  4  1  2  4
           131  7 11  3  7 12 12  5 10  4  3  6  7 11  8  7 18
68 26   0   16  NI I- Z- I  IU U  R  O  *  KU U- NI I- *  CH I'
            38  2  6  2  2  3  2  2  4  1  3  2  2  4  1  2  0
           121  7 11  3  7 12 12  5 10  4  1  6  7 11  8  7 10
```

| | | |

```
 87 31 -86   14  S  A  N. Z- I  IU U  UI I  *  CH I  NI I-
             40  2  5  7  2  2  3  2  2  3  1  2  3  2  4
            125  8 12 14  4  7 12 10  7 10  7  7  9  7 11
 88 32   0    2  I  E
             12  8  4
             39 24 15
```

図A 6 − 3　　時

```
  1  1   0    7  SH I  *  CH I  Z- I
              18  2  2  0  2  4  2  6
              65  8  9  7  7 12  5 17
  2  1   0    7  SH I' *  CH I  Z- I
              16  2  0  0  2  4  2  6
              61  8  5  7  7 12  5 17
  3  2   0    8  H  A  AI *  CH I  Z- I
              19  2  2  0  1  2  4  2  6
              66  6  7  5  7  7 12  5 17
  4  3   0    5  KU U  UI Z- I
              14  1  3  2  2  6
              40  3  9  6  5 17
  5  4   0    7  Z  I  IU U  UI Z- I
              19  2  2  3  2  2  2  6
              68  7 10 12 10  7  5 17
  6  5   0   11  Z  I  IU U  UI I  *  CH I  Z- I
              30  2  2  3  2  2  3  1  2  5  2  6
             104  7 10 12 10  7 10  7  7 12  5 17
  7  6   0    8  Z  I  IU U  NI I- Z- I
              26  2  2  3  2  3  6  2  6
              82  7 10 12 12  7 12  5 17
  8  7   0    9  Z  I  IU U  S  A  N. Z- I
              32  2  2  3  2  4  6  5  2  6
              98  7 10 12 12 10 13  2  5 17
  9  8   0   11  Z  I  IU U  UI I  IO 0  0I Z- I
              22  2  2  3  2  2  0  2  0  1  2  6
              92  7 10 12 10  7  5  7  5  7  5 17
 10  9   0    9  Z  I  IU U  G  0  0I Z- I
              25  2  2  3  2  3  3  2  2  6
              80  7 10 12 12  5  7  5  5 17
 11 10   0   12  Z  I  IU U  R  0  *  KU U  UI Z- I
              29  2  2  3  2  2  4  1  1  2  2  2  6
              97  7 10 12 12  5 10  4  3  7  5  5 17
 12 11   0   11  Z  I  IU U  N  A  N  A  AI Z- I
              30  2  2  3  2  2  4  2  4  1  2  6
              94  2 10 12 12  5  8  5  8  5  5 17
 13 11   0   12  Z  I  IU U  UI SH I  *  CH I  Z- I
              33  2  2  3  2  2  3  2  1  2  6  2  6
             105  7 10 12 10  7  6  5  7  7 12  5 17
 14 11   0   12  Z  I  IU U  UI SH I' *  CH I  Z- I
              30  2  2  3  2  2  3  0  0  2  6  2  6
             104  7 10 12 10  7  6  4  7  7 12  5 17
 15 12   0   12  Z  I  IU U  H  A  AI *  CH I  Z- I
              28  2  2  3  2  2  2  0  1  2  4  2  6
             107  7 10 12 12  6  7  5  7  7 12  5 17
```

```
16 13   0   10  Z  1  IU U  *  KU U  UI Z- I
            23  2  2  3  2  1  1  2  2  2  6
            84  7 10 12 12  6  3  7  5  5 17
17 14   0    9  NI I- Z- I  IU U  UI Z- I
            27  2  6  2  2  3  2  2  2  6
            80  7 11  3  8 12 10  7  5 17
18 15   0   13  NI I- Z- I  IU U  UI I  *  CH I  Z- I
            39  2  6  2  2  3  2  2  3  1  2  6  2  6
           116  7 11  3  8 12 10  7 10  7  7 12  5 17
```

図A 6－4　　枚　数

```
 1  1   0    8  I  *  CH I  M  A  AI I
             22  4  1  2  4  4  4  3  0
             70  8  7  7 12  9 10  7 10
 2  1   0    8  I´ *  CH I  M  A  AI I
             22  4  1  2  4  4  4  3  0
             70  8  7  7 12  9 10  7 10
 3  2   0    6  NI I- M  A  AI I
             15  2  2  4  4  3  0
             55  7 12  9 10  7 10
 4  3   0    6  S  A  M  A  AI I
             20  0  5  8  4  3  0
             70  8 15 20 10  7 10
 5  4   0    7  I  I0 0  M  A  AI I
             21  1  2  3  8  4  3  0
             72  6  7 12 20 10  7 10
 6  5   0    6  G  0  M  A  AI I
             17  3  4  3  4  3  0
             53  5 12  9 10  7 10
 7  6   0    9  R  0  *  KU U- M  A  AI I
             21  2  4  1  1  2  4  4  3  0
             73  8 12  4  3 10  9 10  7 10
 8  7   0    8  N  A  N  A  M  A  AI I
             23  2  4  2  4  4  4  3  0
             74  6 12  6 14  9 10  7 10
 9  8   0   10  H  A  AI *  CH I  M  A  AI I
             21  2  2  0  1  2  3  4  4  3  0
             77  6  7  5  7  7 10  9  9  7 10
10  9   0    8  KI I  IU U  M  A  AI I
             17  1  0  3  2  4  4  3  0
             76  7 10 12 12  9  9  7 10
11 10   0    8  Z  I  IU U  M  A  AI I
             20  1  3  3  2  4  4  3  0
             76  7 10 12 12  9  9  7 10
12 11   0    2  I  E
             12  8  4
             39 24 15
```

図A 6－5　　航 空 会 社

```
 1  1   0   11  NI I- H  0  N- *  KO 0  *  KU U
             38  2  3  2  4  3  1  1 10  1  1 10
            116  7 10  4 15 13  6  4 24  6  3 24
```

```
2  1  0   11  NI I- H  0  N- *  KO 0  *  KU U'
          34  2  3  2  4  3  1  1 10  1  1  6
         110  7 10  4 15 13  6  4 24  6  3 18
3  1  0    5  NI I- *  KO 0
          33  2  4 10  2 15
          68  7 10 20  4 27
4  2  0    8  ZZ E  N  NI I  *  KU U
          34  2  5  4  3  4  5  1 10
          98 10 14  9  8 12 18  3 24
5  2  0    8  ZZ E  N  NI I  *  KU U'
          30  2  5  4  3  4  5  1  6
          92 10 14  9  8 12 18  3 18
6  3  0   19  T  0  A  *  KO 0  *  KU U- N  A  AI I  *  KO 0  *  KU U
          54  1 10  4  1  1  4  1  1  2  2  2  4  0  1  1  7  1  1 10
         161  3 22 14  6  4  8  6  3  5  6  9  8  7  6  3 18  6  3 24
7  3  0   19  T  0  A  *  KO 0  *  KU U- N  A  AI I  *  KO 0  *  KU U'
          48  1 10  4  1  1  4  1  1  2  2  2  4  0  1  1  7  1  1  4
         149  3 22 14  6  4  8  6  3  5  6  9  8  7  6  3 18  6  3 12
8  3  0   13  T  0  A  *  KO 0  *  KU U- N  A  AI I
          33  1 10  4  1  1  4  1  1  2  2  2  4  0
         101  3 22 14  6  4  8  6  3  5  6  9  8  7
9  3  0    3  T  0  A
          18  1 10  7
          53  3 30 20
10 4  0   23  NI I- H  0  N- *  KI I  N- *  KI I  IO 0  OI R- I  *  KO 0
          52  2  3  2  4  3  1  1  2  4  1  1  0  3  0  0  1  4  1  1  8
         182  7  7  4 15 13  6  4 12 15  6  4  3  7  2  4  3 13  7  3 18
              *  KU U
              1  1  8
              6  3 20
11 4  0   23  NI I- H  0  N- *  KI I  N- *  KI I  IO 0  OI R- I  *  KO 0
          48  2  3  2  4  3  1  1  2  4  1  1  0  3  0  0  1  4  1  1  8
         175  7  7  4 15 13  6  4 12 15  6  4  3  7  2  4  3 13  7  3 18
              *  KU U'
              1  1  4
              6  4 12
12 4  0   17  NI I- H  0  N- *  KI I  N- *  KI I  IO 0  OI R- I
          32  2  3  2  4  3  1  1  2  4  1  1  0  3  0  0  1  4
         128  7  7  4 15 13  6  4 12 15  6  4  3  7  2  4  3 16
13 4  0   17  KI I  N- *  KI I  IO 0  OI R- I  *  KO 0  *  KU U
          37  1  2  4  1  1  0  3  0  0  1  4  1  1  8  1  1  8
         130  4 12 15  6  4  3  7  2  4  3 13  7  3 18  6  3 20
14 4  0   17  KI I  N- *  KI I  IO 0  OI R- I  *  KO 0  *  KU U'
          33  1  2  4  1  1  0  3  0  0  1  4  1  1  8  1  1  4
         122  4 12 15  6  4  3  7  2  4  3 13  7  3 18  6  3 12
15 5  0   11  N  A  N. S  E  *  KO 0  *  KU U
          45  2  3  3  4  9  1  1 10  1  1 10
         135  5 15 14 12 20  7  3 20 10  3 26
16 5  0   11  N  A  N. S  E  *  KO 0  *  KU U'
          39  2  3  3  4  9  1  1 10  1  1  4
         121  5 15 14 12 20  7  3 20 10  3 12
17 6  0    2  I  E
          12  8  4
          39 24 15
```

図A 6－6　　空　港

1	1	0	7	S	A	*	P	O	R	O								
			28	1	8	6	O	5	2	6								
			83	8	16	19	1	12	7	20								
2	2	0	11	A	S	A	AI	HI	I	*	KA	A	W	A				
			31	5	3	4	O	2	3	1	1	5	2	5				
			102	12	10	8	4	8	12	7	3	12	6	20				
3	3	0	10	M	E	M	A	M	B	E	*	TS	U					
			37	2	5	2	8	6	1	5	O	3	5					
			101	6	12	7	16	16	5	12	5	10	12					
4	3	0	10	M	E	M	A	M	B	E	*	TS	U′					
			32	2	5	2	8	6	1	5	O	3	O					
			96	6	12	7	16	16	5	12	5	10	7					
5	3	-4	7	M	E	M	A	M	B	E								
			29	2	5	2	8	6	1	5								
			74	6	12	7	16	16	5	12								
6	4	0	9	W	A	*	KA	A	N	A	AI	I						
			28	2	5	7	1	5	2	4	2	O						
			83	5	14	18	3	12	6	10	8	7						
7	5	0	6	KU	U′	SH	I	R	O									
			19	1	2	4	4	2	6									
			66	3	6	15	15	7	20									
8	5	0	6	KU	U	SH	I	R	O									
			19	1	2	4	4	2	6									
			66	3	6	15	15	7	20									
9	6	0	7	O	B	I	HI	I	R	O								
			30	5	3	4	4	4	4	6								
			78	12	4	15	8	12	7	20								
10	7	0	10	H	A	*	KO	O	D	A	*	T	E					
			27	2	5	O	1	4	1	7	1	1	5					
			97	8	12	4	3	16	5	16	7	2	24					
11	8	0	8	A	AI	*	KI	I	*	T	A							
			19	4	O	O	1	5	1	1	7							
			65	9	4	4	5	14	7	2	20							
12	8	0	8	A	AI	*	KI	I′	*	T	A							
			15	5	O	O	1	2	1	1	5							
			58	9	4	4	5	7	7	2	20							
13	9	0	7	A	O	M	O	OI	R-	I								
			17	4	4	2	3	O	1	3								
			82	12	18	8	18	4	6	16								
14	10	0	10	H	A	AI	*	CH	I	N	O	H	E					
			27	2	4	O	1	3	3	3	4	2	5					
			102	8	7	4	6	8	15	8	14	8	24					
15	11	0	10	H	A	N	A	M	A	AI	*	KI	I					
			25	2	3	2	5	2	5	O	O	1	5					
			89	8	10	6	12	7	9	4	4	5	24					
16	11	0	10	H	A	N	A	M	A	AI	*	KI	I′					
			20	2	3	2	5	2	5	O	O	1	O					
			77	8	10	6	12	7	9	4	4	5	12					
17	11	-16	7	H	A	N	A	M	A	AI								
			19	2	3	2	5	2	5	O								
			56	8	10	6	12	7	9	4								

```
18 12   0    8  T  0  0I *  KI I  IO 0
            25  1 10  0  1  3  2  3  5
            76  3 20  7  6  5  4  7 24
19 13   0    6  0  S  A  *  KA A
            32 12  3  8  1  1  7
            83 25 10 16  5  3 24
20 14   0   14  N  A  N- *  KI I  SH I  R  A  H  A  M  A
            40  2  5  4  1  1  4  4  1  1  4  2  4  2  5
           154  6 12 16  6  5 15 15 10  5 12  8 12  8 24
21 15   0    7  NI I- G  A  *  T  A
            28  3  8  2  6  1  1  7
            83  8 24  6 12  7  2 24
22 16   0   10  0  *  KA A  AI I  IA A  M  A
            21  3  1  1  3  2  0  2  3  2  4
            83 12  5  3 10  6  4  6 10  7 20
23 16  22    8  KA A  AI I  IA A  M  A
            17  1  3  2  0  2  3  2  4
            66  3 10  6  4  6 10  7 20
24 17   0    5  0  0I *  KI I
            12  5  0  1  1  5
            41 10  3  7  5 16
25 17   0    5  0  0I *  KI I'
            12  5  0  1  1  5
            36 10  3  7  5 11
26 18   0    7  1  IO 0  N  A  G  0
            25  2  2  3  3  6  3  6
            68  5  8  8  6 15  6 20
27 19   0    5  I  ZZ U  M  0
            23  6  3  4  4  6
            65 15  8 12 10 20
28 20   0    9  T  0  *  KU U' SH I  M  A
            20  1  3  1  1  1  4  1  3  5
            88  3 12  5  3 12 15  8 10 20
29 20   0    9  T  0  *  KU U  SH I  M  A
            20  1  3  1  1  1  4  1  3  5
            88  3 12  5  3 12 15  8 10 20
30 21   0   10  T  A  *  KA A  M  A  *  TS U
            28  1  5  1  1  4  2  6  0  3  5
            90  2 12  5  3 14  7 14  5 10 18
31 21   0   10  T  A  *  KA A  M  A  *  TS U'
            23  1  5  1  1  4  2  6  0  3  0
            82  2 12  5  3 14  7 14  5 10 10
32 21 -31    7  T  A  *  KA A  M  A
            20  1  5  1  1  4  2  6
            57  2 12  5  3 14  7 14
33 22   0    6  KO 0  0I *  CH I
            20  1  9  2  1  3  4
            69  3 24  7  7  8 20
34 22   0    6  KO 0  0I *  CH I'
            20  1  9  2  1  3  4
            63  3 24  7  7  8 14
35 23   0    9  HI I  R  0  0I SH I  M  A
            25  2  3  2  3  2  4  1  3  5
            95  8 10  7 10  4 15  7 10 24
```

```
36 24   0    3  U  B  E
            13  4  1  8
            40 15  5 20
37 25   0   11  M  A  *  TS U  UI I  IA A  M  A
            26  2  4  1  3  1  1  1  3  3  2  5

38 26   0    6  0  0I I  *  T  A
            21  8  2  4  1  1  5
            75 25  4 12  8  2 24
39 27   0    9  F  U  *  KU U  0  *  KA A
            23  2  2  1  1  2  8  1  1  5
            94  6 12  5  3 10 24  7  3 24
40 27   0    9  F  U' *  KU U  0  *  KA A
            21  2  0  1  1  2  8  1  1  5
            87  6  5  5  3 10 24  7  3 24
41 27  40    6  KU U  0  *  KA A
            18  1  2  8  1  1  5
            71  3 10 24  7  3 24
42 28   0   10  M  I  IA A  ZZ A  AI *  KI I
            26  2  4  3  4  3  4  0  0  1  5
            97  6 16  7 15  8 10  4  6  5 20
43 28   0   10  M  I  IA A  ZZ A  AI *  KI I'
            21  2  4  3  4  3  4  0  0  1  0
            87  6 16  7 15  8 10  4  6  5 10
44 28 -43    7  M  I  IA A  ZZ A  AI
            20  2  4  3  4  3  4  0
            66  6 16  7 15  8 10  4
45 29   0    9  KA A  G  0  0I SH I  M  A
            21  1  1  1  3  2  4  1  3  5
            91  3 10  6 12  4 15  7 10 24
46 30   0   11  T  A  N  E  G  A  AI SH I  M  A
            29  1  4  2  4  2  3  0  4  1  3  5
           114  2 12  6 16  6 12  4 15  7 10 24
47 31   0   10  I  IA A  *  KU U' SH I  M  A
            20  0  1  3  1  1  1  4  1  3  5
            94  5  5 12  5  3  8 15  7 10 24
48 31   0   10  I  IA A  *  KU U  SH I  M  A
            20  0  1  3  1  1  1  4  1  3  5
            94  5  5 12  5  3  8 15  7 10 24
49 32   0   11  KI I  *  KA A  AI I  Z- I  M  A
            38  1  5  1  1  6  2  5  2  7  3  5
            99  5 10  5  3 12  5 12  3 14 10 20
50 32   0   11  KI I' *  KA A  AI I  Z- I  M  A
            35  1  2  1  1  6  2  5  2  7  3  5
            94  5  5  5  3 12  5 12  3 14 10 20
51 32  50    8  KA A  AI I  Z- I  M  A
            28  1  4  2  5  1  7  3  5
            83  3 12  5 12  3 14 10 24
52 33   0   12  A  M  A  M  I  I0 0  0I SH I  M  A
            41  4  2  5  2  4  3  6  2  4  1  3  5
           143 12  7 16  6 16  5 20  4 15  8 10 24
53 34   0   12  T  0  *  KU U- N  0  0I SH I  M  A
            27  1  4  1  1  1  3  3  0  4  1  3  5
           104  3 10  5  3  7  8  8  3 15  8 10 24
```

```
54 35   0   12  0  0I *  KI I  N  0  E  R  A  B  U
            28  3  0  0  1  2  3  4  4  1  6  2  2
           122 14  4  4  5 12  8 15 15  5 18  6 16
55 35  54    9 KI I  N  0  E  R  A  B  U
            25  1  2  3  4  4  1  6  2  2
           100  5 12  8 15 15  5 18  6 16
56 35 -54   10  0  0I *  KI I  N  0  E  R  A
            23  3  0  0  1  2  3  3  4  1  6
            93 14  4  4  5 12  8  8 15  5 18
57 36   0   10  I  IA A  M  A  G  A  *  T  A
            27  2  2  2  2  5  2  5  1  1  5
           101  6  7 10  7 16  6 16  7  2 24
58 37   0    7  S  E  N  D  A  AI I
            23  3  5  7  1  4  3  0
            79  8 16 20  5 12  8 10
59 38   0   15  H  A  AI *  CH I  Z- I  IO 0  0I Z- I  M  A
            38  2  4  0  1  3  4  1  2  2  4  1  2  4  3  5
           135  8 12  4  5  8 14  3  8  5 12  4  6 12 10 24
60 39   0    6  0  0I SH I  M  A
            21  8  0  4  1  3  5
            87 24  7 15  7 10 24
61 40   0   12  M  I  IA A  *  KE E  I  Z- I  M  A
            31  3  4  2  3  1  1  3  0  2  4  3  5
           109  6  8  7 12  4  3 16  5  3 15 10 20
62 41   0    8  T  0  0I I  IA A  M  A
            17  1  1  2  1  2  2  3  5
            72  3  5  5  6  7 12 10 24
63 42   0    8  KA A  N  A  ZZ A  W  A
            29  1  3  4  5  4  4  4  4
            80  3 12  6 16  8 12  6 17
64 43   0    7  F  U  *  KU U  UI I
            16  2  2  1  1  2  3  5
            56  6 12  5  3  8  6 16
65 43   0    7  F  U' *  KU U  UI I
            14  2  0  1  1  2  3  5
            49  6  5  5  3  8  6 16
66 43  65    4  KU U  UI I
            11  1  2  3  5
            33  3  8  6 16
67 44   0    8  T  0  *  T  0  0I R- I
            26  1  5 10  1  2  0  1  6
            84  3 16 19  3 10  4  5 24
68 45   0    8  N  A  G  0  0I I  IA A
            21  2  5  2  3  2  1  2  4
            78  6 15  6 12  6  7  6 20
69 46   0   16  KI I  *  T  A  AI *  KI I  IU U  UI SH I  IU U
            32  1  2  1  1  5  0  0  1  1  3  0  2  6  1  3  5
           109  5  5  7  2 12  4  4  5  5  8  4  5 18  4  6 15
70 46   0   16  KI I' *  T  A  AI *  KI I  IU U  UI SH I  IU U
            30  1  0  1  1  5  0  0  1  1  3  0  2  6  1  3  5
           108  5  4  7  2 12  4  4  5  5  8  4  5 18  4  6 15
71 46  70   13  T  A  AI *  KI I  IU U  UI SH I  IU U
            28  1  5  0  0  1  1  3  0  2  6  1  3  5
            92  2 12  4  4  5  5  8  4  5 18  4  6 15
```

```
72 47   0    10  N  A  G  A  S  A  AI *  KI I
             25  2  4  2  5  3  3  0  0  1  5
             95  6 12  6 16 10 12  4  4  5 20
73 47   0    10  N  A  G  A  S  A  AI *  KI I'
             21  2  4  2  5  3  4  0  0  1  0
             81  6 12  6 16 10  8  4  4  5 10
74 47 -73     7  N  A  G  A  S  A  AI
             20  2  4  2  5  3  4  0
             62  6 12  6 16 10  8  4
75 48   0     9  KU U- M  A  M  0  *  T  0
             28  1  2  3  5  4  5  1  1  6
             87  3  6  7 16 10 15  7  3 20
76 49   0     6  F  U  *  KU U  E
             22  2  4  1  1  4 10
             63  6 10  5  3 15 24
77 49   0     6  F  U' *  KU U  E
             18  2  0  1  1  4 10
             60  6  7  5  3 15 24
78 49  77     3  KU U  E
             15  1  4 10
             42  3 15 24
79 50   0     4  S  A  D  0
             16  3  4  3  6
             38  8 11  5 14
80 51   0     4  I  *  KI I
             13  5  1  2  5
             38 12  7  5 14
81 51   0     4  I  *  KI I'
             13  5  1  2  5
             34 12  7  5 10
82 52   0     8  M  0  M  B  E  *  TS U
             30  2  8  6  1  5  0  3  5
             92  8 20 16  5 12  5 10 16
83 52   0     8  M  0  M  B  E  *  TS U'
             30  2  8  6  1  5  0  3  5
             92  8 20 16  5 12  5 10 16
84 52 -83     5  M  0  M  B  F
             22  2  8  6  1  5
             61  8 20 16  5 12
85 53   0    13  N  A  *  KA A  AI SH I  B  E  *  TS U
             32  2  5  0  1  5  0  4  2  1  5  0  3  4
            115  6 12  5  3 12  4 15  7  5 15  5 10 16
86 53   0    13  N  A  *  KA A  AI SH I  B  E  *  TS U'
             32  2  5  0  1  5  0  4  2  1  5  0  3  4
            115  6 12  5  3 12  4 15  7  5 15  5 10 16
87 53 -86    10  N  A  *  KA A  AI SH I  B  E
             25  2  5  0  1  5  0  4  2  1  5
             84  6 12  5  3 12  4 15  7  5 15
88 54   0     9  0  0I *  KI I  N  A  W  A
             23  4  0  0  1  3  2  4  4  5
             90 14  4  4  5 13  6 14  6 24
89 54  88     6  KI I  N  A  W  A
             17  1  3  2  4  2  5
             68  5 13  6 14  6 24
```

```
90 55   0    9  KU U- M  E  I  Z- I  M  A
            31  1  3  4  4  2  2  6  4  5
            79  3  6  6 16  5  3 10 10 20
91 56   0   13  M  I  N  A  M  I  D  A  AI I  *  T  O
            41  2  1  2  5  2  3  1  4  3  4  3  1 10
           122  6  6  6 15  6 12  5 12  5 10 12  3 24
92 57   0    7  M  I  IA A  *  KO O
            18  2  2  3  5  1  1  4
            64  6 10  5 16  8  3 16
93 58   0    6  T  A  R  A  M  A
            19  1  5  1  5  2  5
            66  2 12  5 16  7 24
94 59   0    9  I  SH I  G  A  AI *  KI I
            23  3  4  2  2  4  0  0  1  7
            90 10 15 12  6 10  4  4  5 24
95 59   0    9  I  SH I  G  A  AI *  KI I'
            18  3  4  2  2  4  0  0  1  2
            78 10 15 12  6 10  4  4  5 12
96 59 -95    6  I  SH I  G  A  AI
            15  3  4  2  2  4  0
            57 10 15 12  6 10  4
97 60   0   10  I  IO 0  N  A  *  KU U- NI I-
            23  2  3  1  2  5  1  1  2  3  3
            83  6  7  8  6 16  5  3  8  8 16
98 61   0    7  CH I  *  T  0  S  E
            20  2  3  1  1  5  3  5
            67  8 10  4  3 12 10 20
99 61   0    7  CH I' *  T  0  S  E
            18  3  0  1  1  5  3  5
            62  8  5  4  3 12 10 20
100 61  99   4  T  0  S  E
            14  1  5  3  5
            45  3 12 10 20
101 62  0    7  KO 0  M  A  *  TS U
            23  1  3  4  5  0  3  7
            76  3 11  7 20  5 10 20
102 62  0    7  KO 0  M  A  *  TS U'
            16  1  3  4  5  0  3  0
            66  3 11  7 20  5 10 10
103 62-102   4  KO 0  M  A
            13  1  3  4  5
            41  3 11  7 20
104 63  0    5  0  M  U  R  A
            31 10  6  4  4  7
            70 24 10 10  6 20
105 64  0    4  N  A  H  A
            12  2  3  2  5
            42  6 12  8 16
106 65  0    6  M  I  S  A  W  A
            20  2  3  3  5  2  5
            64  6 10 10 16  6 16
107 66  0    6  R  I  SH I  R- I
            19  1  4  4  4  1  5
            63  5 10 15 15  3 15
```

```
108 67   0    8  0  *  KU U  SH I  R- I
             19  5  1  1  1  4  3  1  3
             83 15  5  3  8 15 16  3 18
109 67   0    8  0  *  KU U' SH I  R- I
             19  5  1  1  1  4  3  1  3
             83 15  5  3  8 15 16  3 18
110 67 109    6 KU U' SH I  R- I
             13  1  1  4  3  1  3
             63  3  8 15 16  3 18
111 68   0    2  I  E
             12  8  4
             39 24 15
```